NON-REDUCTIVE PHYSICALISM,
IRREDUCIBILITY OF THE MENTAL
AND THE PROBLEM OF MENTAL CAUSATION

A STUDY OF DONALD DAVIDSON’S AND GEORG HENRIK VON
WRIGHT’S POSITIONS IN THE PHILOSOPHY OF MIND
This is a time of decline. True. But we have not chosen this time.

Oswald Spengler

If anything about the present century is certain, it is that the power conferred on ‘humanity’ by new technologies will be used to commit atrocious crimes against it.

John Gray

The philosopher will display a kind of anarchy...as a precondition for a new, healthier culture.

Georg Henrik von Wright

It is not the task of philosophy to sing Hallelujah chorus to science or to police its pronouncement.

P.M.S Hacker

...what is the use of studying philosophy if all that it does for you is enable you to talk with some plausibility about some abstruse questions...and if it does not improve your thinking about the questions of everyday life... it is difficult to think well about “certainty”, probability”, “perception” etc. But it is, if possible, still more difficult to think, or try to think, really honestly about your life & other peoples lives.

Ludwig Wittgenstein

The study of philosophy can help make you feel better. It can loosen your mind muscles, smooth out the dogmatic cramps, prescribe exercises which will be a step toward a greater intellectual ease.

Donald Davidson

Irreducibility has its consequences.

David Chalmers
Contents

Acknowledgements........................................................................................................................ 1

Introduction .................................................................................................................................. 2

Chapter one: A challenge for the philosophy of mind and the Davidsonian – von Wrightian response ......................................................................................................................................... 9

1.1 Naturalism and its consequences for the mental aspects of humans................................. 17
   1.1.1 Science gone too far – eliminative materialism ............................................................. 27
1.2 A Davidsonian – von Wrightian response to the challenge ................................................... 35
   1.2.1 A Wittgensteinian view of philosophy .......................................................................... 43
   1.2.2 Davidson’s conception of philosophy ........................................................................... 51
   1.2.3 Von Wright’s conception of philosophy ........................................................................ 68
1.3 A methodological caveat ..................................................................................................... 78
1.4 Conclusions of chapter one.................................................................................................. 79

Chapter two: Non-reductive physicalism and its warrantability .............................................. 80

2.1 One answer to the mind–body problem: Davidson’s anomalous monism ............................. 84
   2.1.1 The place of AM in contemporary philosophy of mind ................................................. 85
   2.1.2 The physicalistic monism of AM .................................................................................. 95
   2.1.3 Davidson’s physicalism .............................................................................................. 106
2.2 Von Wright’s physicalism and its relation to Davidson’s position ...................................... 132
   2.2.1 Von Wright’s supervenience ....................................................................................... 151
2.3 A problem for non-reductive physicalism: Mental properties ............................................. 165
   2.3.1 The nature of mental properties................................................................................... 169
   2.3.2 Why mental properties create a problem for non-reductive physicalism ................. 180
   2.3.3 Davidson on properties ............................................................................................... 189
   2.3.4 Is Davidson’s rejection of properties justified? ............................................................ 198
   2.3.4 Are mental properties physical properties? ................................................................. 205
2.4 A problem for non-reductive physicalism: Token identity .................................................. 219
   2.4.1 A Davidsonian approach: Token identity between the mental and the physical ........ 220
   2.4.2 A von Wrightian approach: Token-identity is a serious confusion ................................ 225
   2.4.3 Token reducibility ....................................................................................................... 229
2.5 The nature of mental phenomena ....................................................................................... 247
   2.5.1 A distinction between conscious and non-conscious beliefs ........................................ 254
   2.5.2 Radical interpretation – Davidson’s conception of the mental ................................... 258
   2.5.3 Interpretation and the nature of mental states .............................................................. 267
   2.5.4 Consequence of interpretationism: Mental phenomena are not states of the brain .... 275
2.6 An alternative to physicalism, dualism and non-reductive physicalism ................................ 282
2.7 Conclusions of chapter two................................................................................................ 299
Chapter three: The irreducibility of the mental ................................................................. 303

3.1 Anomalism of the mental................................................................................................. 315
  3.1.1 Why strict psycho-physical laws are impossible? .................................................. 319
  3.1.2 Why strict psychophysical laws are impossible: A Concrete example ................. 343
3.2 Why the mental has a priority over the physical.......................................................... 364
3.3 The irreducibility of the mental...................................................................................... 389
  3.3.1 What does a reduction of the mental to the physical mean? .................................... 391
    3.3.1.1 New wave reductionism: A challenge for Davidson’s position ......................... 396
  3.3.2 The subjective and intersubjective nature of thoughts .......................................... 419
  3.3.3 The dispositional character of mental concepts ...................................................... 430
    3.3.3.1 AM’s relation to the functional model of reduction ......................................... 436
3.4.4 Externalism ............................................................................................................... 446
3.4 Conclusions of chapter three........................................................................................ 453

Chapter four: The problem of mental causation and the possibility of epiphenomenalism .. 456

4.1 What happens to mental causation if the mental is not an ontological category? .......... 471
  4.1.1 Davidson on causal explanation and causation......................................................... 482
  4.1.2 Davidson on mental causation and rational explanation ......................................... 494
4.2 The new problem of mental causation, epiphenomenalism and precursory reasons ....... 501
  4.2.1 Epiphenomenalism .................................................................................................. 507
    4.2.1.1 Von Wright’s epiphenomenalism ....................................................................... 509
    4.2.1.2 Apparent mental causation ............................................................................... 525
  4.2.2 Precursory reasons .................................................................................................. 531
    4.2.2.1 The consequences of taking precursory reasons seriously ................................ 547
4.3 Conclusions of chapter four.......................................................................................... 551

Appendix: Against Material Mind ..................................................................................... 553

Bibliography ...................................................................................................................... 582
Acknowledgements

This work has been financially supported by the Academy of Finland, University of Helsinki, Finnish Cultural Foundation, Ella and Georg Ehrnrooth Foundation and the Finnish-Norwegian Cultural Foundation. Thank you.

I thank Lauri Kultti for being a friend and an early guide. Sami Pihlström who has given valuable feedback deserves my gratitude. I owe special thanks to Frederick Stoutland who has been an example of a great philosopher and has given me support. I thank and respect Gabriel Sandu, my supervisor, who has endured my pig-headed project.

My parents deserve to be mentioned here since they are part of the causal chain the result of which is me. Eila and Eero, your kindness is invaluable and I thank you dearly.

It is quite likely that without Laura’s emotional support the completion of this thesis would have taken another ten years. Therefore this work is dedicated to her.
**Introduction**

The human being is a social, psycho-physical being. For centuries, religion, philosophy and science have tried to understand and explain, in their differing ways, the nature of this being. As far as we know each human has a physical body. But unlike any other animal, the human is also a self-conscious being with a mental life. Does this mean that humans have a non-physical mind, soul or spirit? Many, especially those with religious beliefs, answer affirmatively. Others, those with a scientific world-view, think that the answer is obviously negative. In philosophical discussions, serious dualists insist that a non-physical soul or mind is an essential aspect of humans, whereas materialists claim that humans consist entirely of matter: the brain is what counts. Both approaches seem to leave something out by ignoring the dual nature of the human being. This being has a physical as well as a mental aspect. But what does it mean to say that a human being is both a physical and a mental being? What, for example, does Ludwig Wittgenstein mean when he writes of his fellow humans that: "My attitude towards him is an attitude towards soul. I am not of the opinion that he has a soul"?¹ How should we understand Wittgenstein’s claim that: “[…] faith is what my heart, my soul needs, not my speculative intelligence. For it is my soul, with its passions, as it were with its flesh and blood, that must be saved, not my abstract mind.”² How should we read Jean-Paul Sartre’s view when he wonders: “I see this man; I apprehend him as an object and at the same time as a man. What does this signify? What do I mean when I assert that this object is a man?”³ What is the key characteristic of a human being, towards whom my attitude is an attitude towards a soul? An answer to this question is far from clear, and at times contemporary philosophy of mind brushes this question aside.

Whereas for a scientist the mind–body problem may have dissolved, a philosopher considers the problem in a more abstract vein, trying to understand, clarify and dispel the conceptual complexities relating to this problem. Scientific discoveries about the brain are continually made; in the United States the years 1990–1999 were labeled as the “Decade of the Brain”. Do these discoveries tell us anything about the soul or about the mind? Are they describing the properties of consciousness? No. The relation of physical and mental reality, the “world knot” as Arthur Schopenhauer it so vividly described it, has not yet been untied. Certain basic philosophical problems about the mind are still without an answer and scientific

---

¹ Wittgenstein, 1953, 178.
² Wittgenstein, 1980a, 33.
³ Sartre, 1965, 341.
knowledge about the mind-brain, although impressive in some areas, is silent with respect to the deepest philosophical problems. The mind–body problem is still a disturbingly difficult question. Philosophers of mind often quote Thomas Huxley, who in the 19th century wrote in amazement: "How it is that anything so remarkable as a state of consciousness comes about as a result of initiating nerve tissue, is just as unaccountable as the appearance of the Djinn, where Aladdin rubbed his lamp in the story…" Some contemporary philosophers have, for different reasons, concluded that the mystery may be unsolvable; they are still as amazed as Huxley was.

In this work my purpose is to consider what kind of perspective on the mind–body problem is taken and can be taken by a philosophical position called non-reductive physicalism. Many positions fall under this label. The form of non-reductive physicalism which I favor is in essential respects the position taken by Donald Davidson and Georg Henrik von Wright. I defend their position and discuss its relationship to the problem of mental causation. In its simplest form, the problem of mental causation is the problem of how mental phenomena like beliefs can cause physical movements of the body. Non-reductive physicalism is committed to the thesis of the irreducibility of the mental. In the philosophical literature, the nature of this irreducibility is expressed in different ways. It is seen as ontological, conceptual, or both. Davidson and von Wright claim that, in some important sense, the mental aspect of a human being does not reduce to the physical aspect, i.e. that there is a gap between these aspects that cannot be closed. Non-reductive physicalism can thus been seen as accepting the view that an attitude towards a human is an attitude towards a soul; the mental aspect (be it a soul, mind, or spirit) is an irreducible part of the human condition. Non-reductive physicalism helps to formulate this claim and explains how it should be understood.

The essential point of non-reductive physicalism – the irreducibility of the mental – and the problem of mental causation are closely related. If mental phenomena do not reduce to causally effective states of the brain, then what justifies the belief that mental phenomena have causal powers? Humans explain their actions in terms of mental causes, and the notion of human agency requires that there are mental causes which are under the control of the acting individual. If mental causes do not reduce to physical causes, then how to tell when – or whether – the mental causes in terms of which actions are explained are actually effective? I will argue that this – how to decide when mental causes really are effective – is the real

4 Huxley, 1866, 193.
problem of mental causation. In the modern philosophy of mind, the problem is formulated in a confusing way. Our current understanding of the problem of mental causation is a result of an “empirical metaphysics” and a consequence of thinking about the nature of philosophy in a certain way. The problem needs to be considered in a broader perspective, one which takes seriously the facts of actual human life. The “solution” to the problem of mental causation that is proposed in this work emphasizes the possibility of epiphenomenalism, but also suggests how this strange conclusion could perhaps be avoided.

Davidson’s contributions to contemporary philosophy of mind are well known; yet it is often claimed that his position of anomalous monism fails as an adequate theory of mind. Although Davidson’s position has faced severe criticism and his specific claims are rejected by most commentators, his attempt to provide a systematic vision of the human mind and its relation to the world has nevertheless been widely acclaimed and he is considered as one of the great philosophers of the 20th century. This is not so clearly the case with von Wright, whose contributions to philosophy of mind are relatively unknown. Although contemporary philosophers may recognize von Wright’s work on logic and some may remember his book *Explanation and Understanding*, it is fair to say that von Wright’s later work has gone largely unnoticed. One important purpose of this work is to clarify some exegetical issues around the works of these two philosophers and thereby dispel the criticism that Davidson’s and von Wright’s views are old-fashioned, relate to an outdated understanding of the nature of philosophy, and are irrelevant to the modern discussion of the mind–body problem. In this work, I describe von Wright’s and Davidson’s place in the contemporary philosophy of mind, discuss their relation to naturalism, and compare their views to the views of other well known modern philosophers. Interesting things about the mind and brain are published constantly. My purpose is to show what kind of relevance von Wright’s and Davidson’s positions might have for some of these views. In doing this I emphasize the similarities between von Wright’s and Davidson’s positions. This has not yet been done in the literature. Insofar as the views of these two philosophers have been compared at all, the focus has been on the differences. The emphasis will be on Davidson’s views, because his work in the philosophy of mind is more extensive than von Wright’s, whose contributions are mainly presented in his last book *In the Shadow of Descartes*.

A great number of articles have been written on Davidson’s philosophy, but the discussion of central aspects of his philosophical views is still ongoing. A good example of

---

5 Perhaps the best general expositions of his philosophy of mind are the somewhat outdated Evnine, 1991, and more up-to-date Joseph, 2001.
this is a recent debate between “world leading authorities on Davidson’s philosophy”: Ernest Lepore and Kirk Ludwig on one side, and Frederick Stoutland, who incidentally knows von Wright’s philosophy very well, on the other. This debate shows that there exist disagreements about the central features of Davidson’s views. My purpose is to clarify Davidson’s position in order to simplify these exegetical debates. I will show that Davidson’s views changed over the years, and thus contend that any interpretation which does not take this into account cannot be a truthful description of his views. It was only at a late stage of the writing process of this thesis that I became aware of two recent studies of Davidson’s philosophy of mind which look quite interesting from the perspective of this work. Neil Campbell’s *Mental Causation: A Nonreductive Approach* and Lars Reinholdtsen’s doctoral thesis *Interpretation and Explanation: Reflections on Donald Davidson’s Philosophy of Mind* contain views which come close to certain conclusions of my own work. However, both these works leave open many questions concerning Davidson’s position, and some of those questions are addressed in this work. A complete clarification cannot be expected, however. Davidson’s views are sometimes inconsistent and leave room for different interpretations. Whereas a lot has been written about Davidson, there is not much of a secondary literature on von Wright’s philosophy of mind. The exegetical issues concerning his philosophy are still largely unresolved, and there is a need for a detailed analysis of his views. The fact that von Wright’s philosophy of mind is unknown to many is unfortunate because, as I will show, his arguments are certainly relevant to the contemporary discussion. The “non-causalist” view often attributed to von Wright and to other “neo-Wittgensteinians” is a relevant contribution to the discussion concerning the problem of mental causation, and I suggest that a solution to the problem should be sought from this direction. Von Wright’s views and arguments are unconventional, both in their style and in their conclusions, from the naturalistic mainstream. His understanding of the nature of philosophy provides an alternative for those who are perhaps frustrated by the naturalistic conception. “What is philosophy?” and “What is a philosophical method?” are important metaphilosophical questions that have great relevance, especially for modern philosophy of mind, which emerges as a hybrid of empirical and conceptual considerations.

It would be a loss for the philosophical community if von Wright’s views came to be forgotten. In addition to von Wright’s published writings, I have therefore studied his unpublished work, hoping to put it in chronological order so as to reveal von Wright’s final

---

conclusions about the mind.\textsuperscript{8} I am focusing especially on von Wright’s views on the philosophy of mind and not so much on his views on the philosophy of action (insofar as these views can be clearly separated). The reason for this is that von Wright’s philosophy of mind has not been much studied, whereas there are interesting analyses about his views on action. At a time when von Wright’s work in the philosophy of mind was largely unpublished, he once wrote that perhaps his purpose in writing about these issues was only to leave a small legacy for his younger philosophical congenial spirits.\textsuperscript{9} Two years before his death, von Wright wrote down that his thoughts were more and more often moving outside the mainstream. At this time, he again expressed his hope that perhaps someday in the future others would dig up some of his later thoughts and develop them further.\textsuperscript{10} My modest attempt here is to cherish this legacy by showing what kind of views von Wright can be seen as defending, and thence suggesting what further explorations, based on these views, might be started in the future. Von Wright’s views can (I believe) provide an interesting perspective on the problem of mental causation, which is a widely discussed contemporary problem in philosophy. This being said, it has to be acknowledged that von Wright’s work was left unfinished and the author himself could not reach a final conclusion about the difficult problems under study. Von Wright was himself acutely aware that further thinking would be required; perhaps congenial spirits would continue his work further.

The main purpose of this work is to clarify the views of Davidson and von Wright and to discuss how the problem of mental causation should be understood from a non-reductive perspective. The substantial, although obviously not conclusive, conclusions that will be presented in this thesis are as follows.

1. The “physicalism” of non-reductive physicalism is unwarranted.

Non-reductive physicalism is meant to be a position which is true to the scientific world-view, according to which everything that exists is in some abstract sense physical. But certain forms of non-reductive physicalism try to maintain also the “common-sense” idea that humans are not merely physical systems whose behavior can be strictly predicted and controlled. Non-reductive physicalism, as a philosophical position, is an attempt to explain how humans can be part of the physical world but also free agents who are capable of choosing and acting out

\textsuperscript{8} I refer to the unpublished writings with UP.
\textsuperscript{9} Von Wright, 1991.
\textsuperscript{10} Von Wright, 2002.
of reasons. A classic formulation of non-reductive physicalism is Davidson’s anomalous monism.\textsuperscript{11} I will consider how well argued the physicalism of AM is, and show that although Davidson and von Wright both accept the view that everything that exists is physical, they do not provide good philosophical arguments in support of this view. My claim is that the physicalistic ontology of non-reductive physicalism is unwarranted and problematic.

2. The mental aspect of humans is irreducible to the physical aspect.

Non-reductive physicalism combines two theses: (a) Everything that exists is physical; (b) Mental phenomena cannot be reduced to states of the brain. I suggest that whereas the argument in favor of physicalism is unconvincing, there are good reasons to believe that mental phenomena are irreducible. Drawing on the views of Davidson and von Wright, my purpose is to explain why mental phenomena as well as our mental vocabulary and mental explanations are all irreducible to their physical counterparts. The irreducibility of the mental is both ontological as well as conceptual. This means that brain research will not solve the mystery of consciousness and that explanations referring to mental phenomena cannot be replaced by explanations which refer to physical phenomena. These assertions support conclusion 1) above, according to which the physicalism of non-reductive physicalism is unwarranted. The truth of physicalism cannot be shown. Mental phenomena and mental explanations are indispensable and will so remain.

3. Non-reductive physicalism cannot solve the problem of mental causation.

If the claim about the irreducibility of the mental is taken seriously, the problem of mental causation cannot be solved. How exactly consciousness can have a causal effect on the physical world remains a mystery. Jaegwon Kim, among others, has repeatedly argued that non-reductive physicalism cannot solve the problem of mental causation. I find his view convincing. My claim is, however, that the failure to solve the problem of mental causation is, pace Kim, a virtue of Davidson’s and von Wright’s accounts – because it leaves room for human freedom. The desire for a reductive solution to the problem of mental causation is a consequence of adopting an overtly metaphysical perspective towards the problem, and the result is a confused way of seeing the nature of the problem.

\textsuperscript{11} AM from now on.
4. A reserved form of epiphenomenalism may be true.

If the problem of mental causation cannot be solved, the possibility arises that epiphenomenalism – a view according to which mental aspect of humans is causally inefficacious and there is no such thing as mental causation – may be true. I claim that there are further reasons, in addition to the failure to solve the problem of mental causation, to think that a reserved form of epiphenomenalism may be true. In making this conclusion, I also show how evidence from neuroscience and psychology tentatively support it.

5. There is a way out of epiphenomenalism.

A view according to which consciousness is always causally inefficacious and freedom a mere illusion cannot be accepted. If the irreducibility of the mental is true there cannot be a reductive solution to the problem of mental causation. I will argue that the only way to reject epiphenomenalism and make sense of mental causation is by rejecting the idea of physicalism. Inspired by the views of Davidson and von Wright, I will show how it is perhaps possible to make sense of mental causation in the context of non-reductive physicalism. A physicalistic answer to the problem of mental causation cannot be given, but a non-reductive answer to this problem is available. This is the only way in which this problem can be “solved”. One therefore has to go beyond conclusion 1) and insist not only that physicalism is unwarranted, but also that the idea of physicalism has to be rejected in order to save human freedom. This could be seen as a kind of transcendental argument: it takes as basic certain facts familiar from ordinary life and asks what kind of ontology must follow. We can, and in fact must, choose how we want to describe human beings – and this choice has consequences for the coherence of a physicalistic world view.

This rejection of physicalism is based on the view that a materialistic metaphysics and its consequent – a material mind – is a threat to the autonomy of human beings. In arguing in favor of the irreducibility of the mental and against physicalism, I want to emphasize the importance of the question of how human nature should be understood, and how humans should be treated. The philosophical problems of the mind should be considered in the broader context of human life and not merely as specific problems waiting for a technical solution in philosophical terms, which are understood only by professional philosophers. Human beings should not be seen as material machines, but as beings with “Wittgensteinian souls”. This is a kind of existential or transcendental choice which, to use von Wright’s
words, must be made but cannot be further grounded. Seeing a human as a mere physical system does not capture the essence of what it is to be a human being. Seeing a human being as a being with soul and not as a machine obviously has far-reaching ethical and moral consequences which, or so I claim, cannot and should not be treated separately from conclusions in the philosophy of mind.

Chapter one: A challenge for the philosophy of mind and the Davidsonian – von Wrightian response

The conclusions of this work center around two closely related problems. The first is the problem of the status of mental phenomena. What is the relation between phenomena like thoughts, desires and wants, which we tend to call “mental”, and phenomena like brain-processes which are understood as being “physical” phenomena? This is the old problem about the relationship between mind and body or mind and matter. Today the problem has turned into the mind–brain problem and is often formulated in terms of consciousness: what is the relation between conscious states, and the brain – or more broadly – physical world? Philosophers talk about the “hard problem of consciousness” and about the “explanatory gap” which exists between the conscious and non-conscious. Is consciousness a “dance of atoms” and can we really understand it as such, or is this kind of claim nonsense? Whatever term – consciousness, mind, soul, or spirit is chosen – a distinction between its referent and physical phenomena is nevertheless a crucial and problematic part of our current self-conception.

The second problem on which this work focuses is the problem of how consciousness, a seemingly non-physical phenomenon, can have causal efficacy in the physical world. Both of these problems have been called “philosophical”, and in the history of philosophy the task of solving them has therefore been a philosopher’s task. But what is a philosophical problem? Is the mind–body problem a philosophical problem? Is the problem of mental causation a philosophical problem? Is there a philosophical and non-philosophical side to these problems? Philosophy can be seen as a systematic attempt to explicate conflicting intuitions. It needs to be considered whether these problems are philosophical, where our intuitions lie, and what kind of clarification of them, if any, philosophy can provide. Do the views of Davidson and von Wright clarify these problems in any way? Sometimes philosophical discussion, instead of clarifying, confuses the views that we might hold if we had not started philosophizing in the first place. It was already Descartes who claimed that the problem of mental causation is
only a philosopher’s problem; in the course of everyday life the problem of mental causation does not arise. With some justification, it could be argued that a notable part of modern philosophy of mind is guilty of mystification. A certain kind of modern philosophy of mind distorts what we already know in order to include our mental reality into a metaphysics of materialism. This philosophy does not consider the lived position of human beings. Instead it uses metaphors when it refers to the “language of thought”, to “brains that make interpretations”, or claims that the “mind is [like] a computer.” But what do these claims really mean? Do they mean, for example, that the mind is a sort of computer? This is not a scientific claim, although it is made by those who want to themselves as “scientific philosophers”. We should however ask: are these metaphors philosophically perplexing or do they cause confusions? Misunderstandings follow when a biased theoretical attitude is taken towards things that are already known. Sometimes explanations are sought when none are needed. For a philosopher the difficult problem is that of deciding what kinds of questions require explanations, i.e. when does one encounter a question which requires philosophical clarification? How to avoid the temptation to search for explanations where none are needed?

The nature of mental phenomena and the question of the possibility of mental causation are problems which are linked to a broader issue, namely that of what our place in the physical reality is. This deep question has dimensions, like the question of the “meaning of life”, that cannot be entirely captured by science but that nevertheless have enormous importance for humans. These questions are not scientific questions, but they are not meaningless questions either. Do we, in recognizing them, have to settle for a view that a person is part of the physical order but also a queer inhabitant of this reality? It is reasonable to claim that, to use J.J.C Smart’s expression, that the question about “our place in the universe” troubles (in one way or another) most humans at some point of their lives.12 With respect to philosophy, it is not an exaggeration to claim, as John Searle has recently done, that “There is exactly one overriding question in the contemporary philosophy... we could put it in this form: How do we fit in?”13 This seems to be the overriding question, at least in the contemporary philosophy of mind.

The importance of this question – how do we fit in? – is profound and should not be ignored when thinking about the mental aspects of human beings. Different belief-systems, like those provided by religions or the ones offered by modern science, give different and often mutually exclusive answers to this question. It is possible to claim that brain-research

---

12 Our Place in the Universe is the title of Smart 1989.
already shows that mind is part of the physical world and that mental causation is a form of physical causation. However, the crucial point is that conscious human beings do not seem to function like mere physical systems, because their behavior cannot be predicted or explained in the same way that the behavior of purely physical systems can. We would be puzzled to hear that somebody could predict our thoughts. This could seem humiliating by downgrading our status as autonomous individuals. As long as this difference between mental and physical systems remains, we can claim that there is a conceptual difference between causal relations involving consciousness and causal relations involving phenomena that are “clearly” physical. Despite the scientific enthusiasm for subsuming everything under the physical, it is of course the case that science has not conclusively shown either that mind is part of the physical world or that mental causation is a form of physical causation. This is why philosophers still fiercely debate these two problems, and the question “how do we fit in?” is such that for many a neutral stand towards it is difficult to take. Searle captures also this aspect perpectively when he writes:

[...] problems of the mind and consciousness are regarded with a passion that is unlike that felt for most other scientific and philosophical issues. The intensity of feeling borders on the religious and the political. It matters desperately to people what sort of solution we get to the[se] problems. Oddly enough, I have encountered more passion from adherents of the computational theory of them mind than from adherents of traditional religious doctrines of the soul. Some computationalists invest an almost religious intensity into their faith that our deepest problems about the mind will have a computational solution.15

We can only guess why this is so; an answer would require considering the sociology of modern philosophy and how it relates to society at large. But there are indeed interesting “ideological” camps in the philosophy of mind – among them “dualists”, “eliminative materialists”, “reductive physicalists”, “non-reductive physicalists”, “emergentists” and so on – not to mention the smaller sub-groups like “pan-psychists” or “epiphenomenalists”. Some of these positions connect concretely to broader issues of a social and political nature.15 I believe that the broader questions, and the relations of philosophical positions to them, should be taken into consideration when discussing the mind–body problem. Science and scientific philosophy can describe how things are, but they cannot answer the question of how things should be. Science aims to provide a clear and coherent picture of reality, but the nature of actual human life is defined by its openness, incoherence and uncertainty. This is part of what it is to be a human being.

15 That this is true, for example, with respect to eliminative materialists is shown in the appendix.
The amount of scientific knowledge at our disposal increases dramatically on a daily basis. Unfortunately, this knowledge has not helped humans to understand the deep complexities of their lives. In many cases, the progress of science has increased these complexities or made them more complicated. The scientific image of man simply does not correspond to the self-image of man which many people have. Humans are still struggling with many of the same psychological problems which have always troubled mankind. Perhaps it could be argued that the reason for this is that these problems are not waiting for a scientific solution, but for philosophical clarification; the nature of the problems should be understood in the right way. It is very unfortunate that many philosophers are not concerned about the nature of their task, about the nature of philosophy, or about the question of whether modern philosophy could be unhealthy in some ways. Partly because of this, philosophy is distancing itself from the actual lives of people. In times like ours when trust in “progress” and in “rational” solutions has taken over, philosophers who want to stay in their ivory towers are ignoring problems which truly trouble humans. Perhaps it should be required that especially in our times, when certain aspects of progress threaten the very existence of humans, philosophizing as an activity should have concrete consequences. It should aim at being an activity which helps people in some way and reveals problematic aspects of our era. One important merit of von Wright’s work is that, as we will see in the final chapter, he was aware of the mood of our times and did not hesitate to express skepticism towards it.

A modern person has reasons to think that human beings, or at least their bodies, are physical objects. Being physical objects, humans obey the laws of nature; their bodies cannot escape the realm of physical laws. If current scientific world view depicts correctly the origin of humans, it is then quite clear that we are part of the long chain of evolution and nothing supernatural, like divine intervention, belongs in the story concerning the history and nature of humans. But it seems that humans have also free will; we can choose what to do with our physical bodies. The claim that human as a whole obeys the laws of nature is therefore rendered problematic: there are no laws of nature – that we know of – which apply to the mental realm. Whereas the story of our evolution as a species is known quite well, the question of how the kind of freedom that humans enjoy, or suffer from, is possible has not, of course, pace Daniel Dennett and others, been explained naturalistically. The dual nature of humans – the problem of how to reconcile free will with the fact that humans are part of the physical order – is, as the histories of philosophy and religion testify, one reason which can

16 For Dennett’s recent approach see his *Freedom Evolves*. 
lead to the postulation of a mind or soul. The mind, soul or spirit must be something other than a physical entity if it is to avoid the grasp of physical laws. Although the triumph of materialistic metaphysics is evident today, our self-conception is, for better or worse, nevertheless still essentially dualistic. This does not apply just to the views of laymen but to scientific talk as well. Doctors talk of mental diseases and wonder how to improve the mental and physical health of their patients. It may be the case that mind is nothing “over and above” the brain but if this is so, people have not really grasped what this means.

In the past, the mind–body problem and the problem of mental causation have been the concern of philosophers. However, it is not clear any more, if it ever was, what a philosophical problem is. Thus the nature of philosopher’s task is also unclear. The nature of philosophy and the task of the philosopher are questions which have troubled philosophy and its self-image for a long time, but during the 20th century they have become even more pressing. This situation is a result of the naturalization of philosophy; naturalists claim that philosophy should be continuous with science, that there is no special philosophical method, that there are no philosophical problems, and that the truths of philosophy do not enjoy a privileged status. Philosophy has no special way to discover truths about the world, and the correct method for the pursuit of truth is the scientific, empirical method. W.V.O. Quine, the teacher of Davidson and an important figure behind contemporary naturalism writes: “We naturalists say that science is the highest path to truth…”17 and therefore: “Naturalism looks only to natural science… for an account of what there is and what there is does.”18 Jerry Fodor, a famous contemporary naturalist, claims that: “Science discovers essences…”19, i.e. it is science and only science which describes how things really are. The status of naturalism and the justification for it are under heated debate.20 For the purposes of this work, it is important to ask whether the naturalistic challenge renders obsolete the need for philosophical investigations. What is the relevance of Davidson’s and von Wright’s philosophical arguments if the naturalistic philosophers are correct? What is the relevance of conclusions which one has arrived at by using a philosophical method?

---

17 Quine, 1995, 261.
18 Quine, 1992, 9. Quine’s impact on contemporary naturalism is widely recognized. His importance for eliminative materialism, the most extreme version of contemporary naturalism, is also often acknowledged.
19 Fodor, 1998, 5. Compare, though: “I don’t know whether science discovers essences. It may be that philosophers make them up.” (Fodor, 1987, 49)
The Quinean view, which is accepted among many modern naturalists, is that scientific claims have a “higher status” than other claims, and only they express truths about reality. The knowledge that traditional philosophy produces is “factually empty” or “conventional” as, for example, “neurophilosopher” Paul Churchland claims.\textsuperscript{21} If this is so, a conflict between philosophical claims and scientific claims is evident at some point. The question about the “threat of science” is relevant for the purposes of this work, because Davidson’s and von Wright’s philosophical views, or philosophies, are especially vulnerable to the naturalistic challenge. The relevance of Davidson’s and von Wright’s views for the modern discussion depends in part on the question of whether or not philosophy can or should make a contribution that is something other than the contribution made by science. In my defense of the Davidsonian–von Wrightian conception of the philosophy of mind, I try to show the importance of genuine philosophy, as opposed to science, for the psycho-physical problem.

Although the mind–body problem and other problems related to it have been discussed for centuries, the approach called “philosophy of mind” is new.\textsuperscript{22} If philosophy is understood as \textit{conceptual analysis}, which in the case of the mind aims to clarify the nature of mental \textit{concepts}, then the contemporary naturalistic challenge against it is obvious. The challenge is posed by science and its ability to answer questions that have been thought to be philosophical, and perhaps unsolvable. The empirical method has proved to be successful with respect to the traditional questions, whereas conceptual analysis has produced only conventional knowledge. Whereas in previous centuries the question about the nature of the non-physical mind was a perfectly legitimate philosophical question, a view according to which humans have such a mind is thought to be naïve and “unscientific” today.\textsuperscript{23} Science has thus solved, at least to its own satisfaction, a problem which was philosophical in the past. Currently mind, or consciousness, is equated with the brain by most naturalistic philosophers, and therefore questions about the mind become questions about the brain. The view according to which study of the brain is also straightforwardly study of the mind is tempting. Patricia Churchland, a philosopher and a neuroscientist, i.e. a neurophilosopher, notes: ”In a way, nothing is more obvious than that philosophers of mind could profit from knowing at least something of what there is to know about how the brain works. After all, one might say, how could the empirical facts about the nervous system fail to be relevant in the philosophy of

\textsuperscript{21} Churchland, 2007a.
\textsuperscript{22} According to Richard Rorty (1994) the term came into currency in the 1950s.
\textsuperscript{23} For a discussion how “unscientific” philosophical positions are currently treated, see Meixner, 2005.
mind.” The neurophilosopher’s view is straightforward; it is obvious that philosophers of mind benefit from the findings of neuroscience. How could it be otherwise? How indeed?

One cannot fail to appreciate how fast brain research is progressing. Interesting facts about the brain are revealed constantly, and what sounded like science-fiction or was almost unimaginable just a few decades ago is reality now. Perhaps even a committed dualist has to pay some attention to these results. Many of the problems which are intuitively identified as problems of the mind – for example questions about personality – are being solved, or so it is at least being claimed, by studying the brain. If one is not a “new-mysterian”, the scientific view about the mind-brain holds the promise that all problems of the mind will receive, in principle at least, an empirical answer. It could be claimed that in this context the need for conceptual analyses disappears. What could they reveal? Science has already explained how things are, why should anyone bother with an obscure analysis of concepts which may not even touch reality at its joints.

Is there any place for philosophy of mind in this kind of intellectual atmosphere? If the answer is affirmative, then the crucial question is that of how a philosophical investigation of the mind differs from a scientific inquiry. My claim is that Davidson’s and von Wright’s positions are especially vulnerable to the challenge that science poses for the philosophy of mind. Because I take their views to be relevant for the contemporary philosophy of mind, my purpose is to show how a position influenced by their views could perhaps overcome the challenge. This, in turn, is an attempt to show that there is a place and a need for a philosophy of mind which is clearly non-scientific and non-empirical. The claim is that the contribution of philosophy is something other than the contribution of science. This may sound obvious; yet the view is challenged by many naturalists. Even if one agrees with the naturalists and acknowledges that “of course” the contribution of philosophy differs in some way from that of science, it should be noted that it is not obvious how the contribution of philosophy differs from the contribution of science and what the consequences of this kind of philosophy should or could be. Professional philosophers are often surprisingly indecisive with respect to these questions; they do not consider the nature of their task, even though there are “ideological camps” among philosophers, and although philosophy may have consequences which are political and ethical. Any philosophy which is a result of accepting a certain ideology is superficial, because it leaves open the basic question; what is it that philosophers are doing?

25 For a different view see McGinn, 1994.
The view of philosophy of mind as conceptual analysis is not one that is favored today. One reason for this is the rise of empirical philosophy of mind, which is influenced by and integrated with cognitive science, neuroscience or to some other research line of natural or physical science. With such empirical influence comes the question of how the discipline called philosophy of mind should be defined. What counts as “legitimate” philosophy of mind, and who are its practitioners? There clearly exists an empirical trend, but philosophers do not agree on the question of what philosophy of mind is and how it should be pursued. The lack of agreement on these questions is what makes them pressing and interesting. What is at stake is the nature of the proper philosophical method. Naturalists claim that the proper method is ultimately empirical research. Some of those who reject this view think that the proper method is the analysis of concepts. I believe that a particular view on the philosophy of mind, such as Davidson’s or von Wright’s, will not be correctly evaluated if there is no agreement on the question of what it is that a philosopher of mind should do. If, as some would argue, the purpose of philosophy of mind is to provide theories which explain in scientific terms, for example, what the relation between mind and brain is, then an approach which focuses on the analysis mental concepts is automatically unsatisfactory. If, as some others would argue, the purpose of philosophy of mind is to clarify problems through conceptual analysis, then an approach which focuses on discovering new empirical facts is unsatisfying.

The “philosophy of mind as conceptual analysis” and the “philosophy of mind as empirical research” approaches are so distinct from one another that even the productive results of one approach are often ignored or downgraded by the other. The contemptuous attitude which certain scientists or scientifically oriented philosophers have towards more traditionally oriented philosophers shows how deep the disagreement is. In the next section, some of the consequences of scientific philosophy of mind are considered and found problematic. The discussion could be called “metaphilosophical” because it includes both describing views of modern philosophers and considering what the motivation for these views is. Whether this kind of analysis is relevant for the substantial philosophical issues is partly a matter of intuition. I claim, for the reasons that follow, that such an analysis is relevant.
1.1 Naturalism and its consequences for the mental aspects of humans

Modern science challenges philosophy, the “mother of all sciences”. Jaegwon Kim has noted that the development of cognitive science and other sciences studying the brain has “[…] to some extent (some will say fundamentally), changed the character of philosophy of mind….”

When faced with this kind of view, a question worth asking is: “What has changed?” Is, or was, there a clearly defined enterprise called “philosophy of mind”, the essence of which changed because of the development of science? What was this pure philosophy of mind which has changed? Perhaps the kind of conceptual analysis pursued for example by Ryle and Strawson? Yet, if one agrees with Rorty that philosophy of mind is whatever philosophers of mind think they are doing, then there can be different ways to practice “philosophy of mind” and these ways can exist in their own right. This would mean that the character of philosophy of mind has not really changed, but it has attracted competition from cognitive and other sciences which, according to some philosophers, perhaps undermine the importance of philosophy of mind. Kim, for example, thinks that there is no real connection between the kind of work that he does in the philosophy of mind and the results of the sciences of cognition or consciousness. He is thus among those whose philosophical views have not really changed. To give a counterexample, Ned Block argues that philosophy of mind, as opposed to other branches of philosophy, is particularly amenable to empirical approaches and that good philosophy of mind always has some empirical grounding. There are thus foundational disagreements that influence the question of what counts or should count as legitimate, interesting or good philosophy of mind. Contemporary philosophy of mind relies heavily on empirical research methods and on empirical results. Von Wright once claimed that in modern philosophy Russell had defeated Wittgenstein, by which he meant that the empirical method had overtaken conceptual analysis. Perhaps this general trend is what partly explains why the character of philosophy of mind has changed in the way Kim suggests.

What weight does the term philosophy carry in the kind of philosophy of mind which is highly empirical? Not much, because philosophy of mind which is a result of a naturalistic view includes the idea that although philosophical problems are very general and very broad, there is nevertheless no difference in kind between these problems and problems that are

---

26 Kim, 1996, xii.
clearly empirical and amenable to scientific solutions. According to this view, there are no real philosophical problems, only problems whose connection to observation-based solutions are more or less distant. Questions that cannot be solved through empirical methods are not interesting because an analysis of them would amount to mere *speculation* or to “playing with words”. Russell’s long shadow influences the philosophical climate of the 21st century. The problem of philosophy, according to many naturalists, is that it has not been empirical enough. In the history of philosophy up to the present day, there has been too much thinking but not enough experimentation. The belief is that *a priori* solutions are flawed and the old-fashioned approach is not going to lead anywhere because it does not provide “real knowledge” about the world. This is what a naturalist claims. She criticizes philosophers for the fact that they have been carrying out conceptual research from the armchair, from which one can get only a twisted perspective on reality. Certain modern philosophers never tire of ridiculing Descartes because he thought that the interaction between mind and body took place in the pineal gland. But was Descartes a bad *philosopher* and was he doing bad *philosophy*? The same question can be asked with respect to Wittgenstein, Nietzsche and Sartre, just to give a few more examples. A naturalist would no doubt claim that in some sense philosophy done by traditional philosophers is of a lower quality than current naturalistic philosophy.

In her book *Neurophilosophy*, Patricia Churchland advises philosophers that if they want to get out from the “narrow canyons of the commonsense conception of the world” they should welcome the alliance of philosophy and the hard sciences.29 The most desirable result is nevertheless not an alliance of philosophy and hard science; it is the replacement of the former by the latter. Looking back, it is easy to see, according to the “neurophilosophers”, that non-empirical philosophy of mind has been making little or no progress. The uselessness of philosophy is sometimes emphasized also by notable scientists. Steven Weinberg, a Nobel laureate in physics, has noted that: “I know of *no one* who has participated actively in the advance of physics in the post-war period whose research has been significantly helped by the work of philosophers.”30 Among naturalists, the current view seems to be that philosophers of mind, instead of helping scientists, are merely hindering progress. Neuroscience and its progress look promising by contrast. Science can provide solutions to problems which have

29 Churchland, 1988, 3. Paul Churchland shares this view: "The single most important development in the philosophy of mind during the past forty years has been the emerging influence of the philosophy of science" (Churchland, 1989, xii). According to those who share Churchland’s vision (e.g. Bickle et al. 2006) *Neurophilosophy* is considered to be a hallmark of neuroscientific philosophy, since it made (some) philosophers finally realize the relevance of neuroscientific findings for the philosophy of mind.

30 Weinberg, 1993, 133-134.
been considered philosophical in the past, and it thus has potential to replace the more philosophical approaches. Churchland writes:

It is difficult to resist the excitement that now typifies so much research in the neurosciences...The excitement is generated in part because neuroscience is science, and in pushing back the bounds of darkness it is discovering surprising new things...  

What is exciting, according to Churchland and other neurophilosophers, is that the mind is finally amenable to scientific study. Those philosophers who expect that neuroscience will have only a minor impact on philosophy are making a bad judgment error. Among certain philosophers, the possibility of the scientific study of the mind has led to the conclusion that such scientific study is the only alternative when it comes to the question of how the problems of the mind should be studied. Scientific study of the mind has thus become not only an alternative to philosophy but a mandatory approach, which is superior when compared to other ways of comprehending the world. One is reminded of Russell’s view according to which work in philosophy should eventually come to resemble work in physics. The challenge of science is a concrete challenge for philosophy, because it can be imagined that future philosophers will be advised to focus increasingly on the empirical results of science. This would change the nature of philosophical investigations as they are traditionally understood. What would be the philosopher’s task in this kind of setting?

The methodological import of naturalism, according to which science is the highest path to truth, and its consequential ontology – physicalism – has in some philosophical circles led to the conclusion that if mental phenomena are real they must be “naturalizable”. Some naturalists, especially eliminative materialists, claim that if mental phenomena like beliefs and thoughts cannot be naturalized then there are no such phenomena. The elimination of these mental phenomena is a dramatic step given the role that these phenomena have in human life, and this elimination goes against our common-sense views about the status of mental phenomena. I believe we can imagine a situation in which an eliminative approach might compete with some other view about the nature of human cognition. Which kind of position is chosen could have some very practical consequences. Some of these consequences will be considered in the last chapter, where I claim that we should not defend a purely physicalistic conception of human beings as long as we want to avoid the consequences that this may bring. We can instead choose the ontology of everyday life as basic and consider the prospects of physicalism from this point of view.

---

31 Churchland, 1988, 10.
Many are not willing to accept the common sense ontology as basic. According to naturalists, “intentional irrealism” follows if the “proper place” of mental phenomena in the physical world cannot be shown. Stephen Schiffer claims that “[…] we should not be prepared to maintain that there are… psychological facts unless we are prepared to maintain that such facts are completely determined by, and nothing over and above, physical facts.”

This view does not yet methodologically privilege natural science in questions about the mind, but it asserts that a discussion of psychological facts should include the view that these facts are dependent on physical facts. Fodor is more explicit when he writes: “[…] the deepest motivation for intentional irrealism derives… from a certain ontological intuition: that there is no place for intentional categories in a physicalistic view of the world; that the intentional can’t be naturalized.”

The physicalistic world view, expressed by Schiffer, is the basis for intentional irrealism if a naturalization of mental phenomena is required for their vindication. That naturalization of mental phenomena is required for their vindication is explicitly expressed by Fodor: “Psychologists have no right to assume that there are intentional states unless they can provide… naturalistic sufficient conditions for something to be in an intentional state.”

A psychologist must be able to provide “naturalistic conditions” for intentional states. Here a metaphysical principle, “an ontological intuition”, the source of which is unclear, allegedly guides what psychologists have or do not have the right to assume. In the contemporary philosophy of mind the requirement of naturalization, which is an abstract metaphysical principle, is thus taken very seriously. Fodor writes:

Either there is a naturalistic theory of representation – in which case, it is the solution to the problem of intentionality – or there is no naturalistic theory of representation, in which case, I for one, will give it all up, become an eliminativist about the mental, and opt for early retirement.

---

33 Schiffer, 1982, 119.
34 Fodor, 1987, 97. Cf. “… there is a worry that the semantic (and / or the intentional) will prove permanently recalcitrant to integration in the natural order” (Fodor, 1984, 232). A closely related reason why mental irrealism could turn out to be true lies in the possibility that the causal efficacy of mental phenomena cannot be established. This is the fear of epiphenomenalism. Epiphenomenal phenomena that have no causal efficacy can be considered unreal. Fred Dretske (1989, 1) notes: “If beliefs and desires are not causally relevant to behavior, I… fail to see why it would be worth having them… If reasons aren’t causes, one of the chief – indeed (for certain people) the only – motive for including them in one’s inventory of the mind, vanishes.” A similar worry, which interestingly has just the opposite source, is the belief that the success of physicalism – i.e. showing how reasons are causes will render reasons qua reasons causally inert. The methodological demand of naturalism has the consequence that science is privileged. If a science of mentality is impossible, then mental phenomena are less than real. According to Fodor (1994a) and Kim (1984), laws of mentality are required for a science of mentality and therefore the possible lawlessness of the mental, defended by Davidson, also implies the irrealism of the mental.
35 Fodor, 1994, 5.
36 Fodor, 1998, 185. So either there has to be a scientific theory of mental representations or eliminativism is true. This can be compared to the view of Stephen Stich, who argued for eliminativism but later changed his mind stating that: “It is, I think, deeply irresponsible to suggest that research be stopped and laboratories closed because the work being carried on there fails to live up to some vaguely stated ontological standard whose
The naturalist claims that a philosopher who does not want fall back on dualism has only two options with respect to mental phenomena: they must be reduced to non-mental phenomena or eliminated. The status of mental phenomena is clear if the approach of a naturalistic philosophy of mind is chosen.

Less clear is how seriously the principle of “reduce or be eliminated” can be taken. Psychologists, sociologists or laymen are not going to stop referring to mental phenomena via mental concepts while patiently waiting for a naturalist to justify the use of this conceptual framework and its consequent ontology of the mental. In its demand that naturalizations be made necessary for the vindication of mentality, naturalistic philosophy suggests that facts already certain from everyday experience require an even firmer basis. Science has become “a transcendental ground”, which must ultimately secure facts already familiar to all. But no one, not even Fodor, is really going to turn eliminativist if it can be shown that mentality will not smoothly reduce to the “natural order”. If, on the other hand, all the talk about intentional irrealism or the non-existence of propositional attitudes is just polemical, one can ask what the point of this talk is. If naturalists are not really claiming that the results of science are privileged, then what are they claiming? If they are claiming that science is truly privileged, then they should really think that intentional irrealism is a viable option. They should really think that common-sense talk about mental phenomena has a second-grade status. Yet, to my knowledge, there are no philosophers who take this possibility seriously. There are no philosophers whose lives are actually governed by this principle. A question worth asking is whether philosophers should “live according to their views”; perhaps this is too much to ask. One can only hope that philosophers would aim for intellectual honesty and that they would take philosophy seriously. Perhaps good philosophy should really change the way we think and act.

The general naturalistic challenge to philosophy is that philosophy is rendered useless because all interesting questions are considered to be amenable to scientific treatment, and the scientific method is therefore privileged. The challenge for the philosophy of mind is the claim that mental phenomena need “scientific vindication” through reduction. Without such vindication, mental phenomena may turn out to be unreal. The challenge of this claim is that importance has not been made clear.” (Stich, 1996, 150) Fodor’s ontological intuition seems to be Stich’s ontological standard, which should not constrain research.

37 It is not clear what “scientific method” includes or excludes. What counts as acceptable science is a question that will not be considered here. It is noteworthy, though, that Fodor, for example, claims that psychologists, who should count as scientists, have “no right” to assume that there are psychological states unless they can show, in “non-psychological” i.e. in “naturalistic” terms, how such states are possible. This suggests that the vocabularies and methods of the natural sciences should be privileged. But why should a metaphysical principle, an “ontological intuition”, guide what can be “assumed”?
the study of mental concepts which are familiar to all is useless, because these concepts fail to refer to real phenomena. The suggestion that mental phenomena may turn out to be non-existent is a queer consequence of the view according to which science is the highest path to truth. Fodor, among others, claims that the world picture given by natural sciences has absolute priority, and this being the case: “[...] the philosophical problems about mind and world have to be situated within the general scientific enterprise, if literal truth is what philosophers aim for.”

Fodor is eager to emphasize that there are “ways of talking” and then there is science with its literal truths. Only science tells the truth about the world. The scientific vocabulary is thus a kind of super-vocabulary because it allegedly describes the world as it really is.

The claim that scientific talk has a priority in questions of ontology is the basic view of the post-Quinean reductive and eliminative materialists. What kind of status is given to scientific talk and what role is granted for science? These are crucial questions dividing contemporary philosophers of mind. Whereas Fodor’s naturalism is based on the view that science discovers essences and that “literal truths” can be expressed only in the scientific vocabulary, many see the commitments and consequences of naturalism differently. It is not

---

39 Consequently, finding a place for mentality requires “[...] showing how you could have... a respectable science whose ontology explicitly acknowledges states that exhibit the sorts of properties that commonsense attributes to the attitudes” (Fodor, 1987, 10). As noted already, it is unclear what counts as “respectable science” and whether the distinctions between sciences could be anything more than arbitrary distinctions based, for example, on sociological factors.

40 Philosophers have not agreed what “naturalism” exactly means and the discussions of its status are ongoing. We can agree with Charles Taliaferro’s (2000, 134) remark that: “Sometimes naturalism is simply another name for a thoroughgoing form of physicalism, while at other times naturalism is simply any view of the world that is incompatible with supernaturalism.” In a critical discussion of naturalism Craig and Moreland (2000, xi) note: “[...] naturalism usually includes… (2) a Grand Story which amounts to an etiological account of how all entities whatsoever have come to be told in terms of an event causal story described in natural scientific terms with a central role given to the atomic theory of matter and evolutionary biology; and (3) a general ontology in which the only entities allowed are those that bear a relevant similarity to those thought to characterize a completed form of physics.” According to Quine (1993, 1, my emphasis): “[...] physicalism is bound to have... important side effects in the framing of more special hypotheses in various branches of science; for physicalism puts a premium on hypotheses favorable to closer integration with physics itself.” An important remark about the nature of naturalism is made by Wagner and Warner (1993, 1) when they claim that “...we take naturalism to be the view that only natural science deserves full and unqualified credence.” Warner (1993, 212) repeats this in noting: “Naturalism... takes natural science as a paradigm of justified belief. The idea... is that only scientific beliefs are legitimate or that these have more legitimacy than any others.” As we have seen, contemporary naturalistic philosophy of mind subscribes to these doctrines. In a recent critical study of naturalism, De Caro and Macarthur (2004) separate three influential types of naturalism. They are: 1) Ontological scientific naturalism, which holds that the entities posited by acceptable scientific explanations are the only genuine entities there are; 2) Methodological (or epistemological scientific naturalism), which holds that it is only by following the methods of the natural sciences that one arrives at genuine knowledge; 3) Semantic scientific naturalism, which holds that the concepts employed by the natural sciences are the only genuine concepts we have, and other concepts can only be retained if we can find an interpretation of them in scientifically respectable terms. De Caro and Macarthur (2004, 7, 9) note that: “[...] scientific naturalism tends toward a global doctrine, committed to all of these versions [1-3] together, on the basis of aspiration for a complete and exhaustive
clear what a proper naturalist can or must accept. There is thus, so I would claim, an intrinsic tension in modern naturalism. For example, according to Rorty, to be a naturalist:


An interesting difference between Rorty’s and Fodor’s versions of naturalism can be seen. According to the former, naturalism goes well with anti-essentialism, whereas according to the latter naturalism goes together with essentialism. To acknowledge that naturalism can be, and is, understood in different ways is to emphasize that a philosopher can be a “naturalist” without being committed to the doctrines that “science is the highest way to truth”, that “only natural science tells what there is or what there is does”, or that “there are essences to be discovered”. The form of naturalism which accepts these doctrines sees other positions in pejorative terms – as an “unscientific” or “supernatural” sliding towards dualism, which is the worst position that a modern philosopher of mind can defend. But the criteria of good philosophy may differ from the criteria of good science, and to dismiss philosophical views simply on the grounds that they are “unscientific” is to slide towards a form of scientism, which (or so I would claim) is a dangerous position. Davidson and von Wright can be counted among naturalists as long as the commitments of naturalism are not understood in the way they usually are in the post-Quinean philosophical atmosphere. This means that whereas Davidson and von Wright are – or can be seen as – naturalists, they are certainly not the kind of “naturalizers” of philosophy that Quine, Fodor, the Churchlands and other truly physicalist philosophers of mind are. The basis for this difference goes back to differing views about the nature of philosophy itself. It could be said that the latter naturalists, following the logical positivists and for example Russell, want to naturalize philosophy, whereas some philosophers following for example Wittgenstein want to resist this tendency.
What naturalistic alternatives are there to Quinean naturalism? According to Rorty’s naturalism: “[…] we should let a thousand vocabularies bloom and then see which survive.”

Scientific vocabularies are not given an automatic privilege because they claim to describe the true essence of reality. They are not privileged for ontological reasons. If they are privileged, it is because they perform certain tasks quite well. But this is really no reason to give these vocabularies a special status, because other vocabularies perform well certain other tasks which cannot be carried out by the scientific vocabularies. The pragmatism in Rorty’s naturalism becomes evident, especially in his later views. A position which, by rejecting “supernatural” entities, is still a version of naturalism but which stands in a sharp contrast to reductive version of naturalism is expressed when Rorty writes that:

[…] all our idioms are tools for coping with the world. This means that there can be no philosophical interest in reducing one idiom to another…. As pragmatists see it, we are equally in touch with reality when we describe a hunk of space-time in atomic, molecular, cellular, physiological, behavioral, intentional, political or religious terms. Looking for an ontological or epistemological gap between such idioms strikes pragmatists as like looking for such gaps between a small Phillips screwdriver and a large crescent wrench; there are all sorts of similarities and differences, but none of them have ontological or epistemological import. There is no invidious ontological or epistemological distinction to be made, for example, between physics and literary criticism.

There is no philosophical interest in reducing certain vocabularies to other vocabularies. This is something accepted also by Davidson and von Wright. As we will see, Rorty’s overall understanding of naturalism comes very close to Davidson’s and von Wright’s positions. Contemporary versions of reductive naturalism are strikingly different from “classical naturalism”, of which the most famous modern proponents are the American pragmatists. Why naturalistic philosophy has taken a distinctively scientific turn is an important and a

---

43 Rorty, 1970, 182. The pragmatic version of naturalism can be compared to Quine’s naturalistic view. According to Quine, mental phenomena do not belong to the scientific conception of the world, there are no facts of the matter about them, and therefore they may have to be eliminated. We are told: “The issue is… whether, in an ideal last accounting of everything or a present practical accounting of everything we can, it is efficacious so to frame our conceptual scheme as to mark out a range of entities or units of a so-called mental kind in addition to physical ones. My hypothesis, put forward in the spirit of a hypothesis of natural science, is that it is not efficacious.” (Quine, 1953, 214) In a famous passage Quine (1960, 221) writes: “One may accept the Brentano thesis either as showing the indispensability of intentional idioms and the importance of an autonomous science of intention, or as showing the baselessness of intentional idioms and the emptiness of a science of intention. My attitude, unlike Brentano’s, is the second.” These views should be compared to Quine’s later suggestion, which has a resemblance to pragmatism: “There is a scope for science on the intensional side too…. Both [extensional and intensional ways of accounting for all events] count as science – good, bad, or indifferent…. We must in any event settle for multiple scientific theories, jointly true.” (Quine, 1990, 72-73) This is however not the kind of pragmatism which Rorty (1994, 126), for example, strives for when he says that: ”A pragmatist should not say.. .that physics tells us all we need to know about the way the world is, but rather should follow Goodman and Putnam in saying that there is no ‘Way the World Is’”. Whereas Quine’s view gives a privilege to physics in dictating what there is, his later doubts about the meaningfulness of the question of “what there really is” moves his position closer to the kind which Rorty suggests.

very interesting question. Perhaps this turn is just a result of the “mood of our times”, which von Wright perspicuously describes thus:

The enormously enlarged access to factual information thanks to electronic technology, in combination with spectacular advances particularly in the experimental sciences and their technological applications, has resulted in the hypertrophy of knowledge at the expense of understanding. This tends to encourage a corrupted paideia, an unwholesome panscientist atmosphere or implicit belief that increase in knowledge alone can successfully cope with all problems worthy of rational consideration in the moral and societal sphere.\textsuperscript{45}

As was briefly suggested before, the increase in scientific knowledge has not led people to solve certain fundamental problems of life; on the contrary, sometimes scientific progress has raised these very problems. In philosophy Russell has won, but the consequences for philosophy have been unfortunate. Scientific philosophy is not really science but the kind of panscience of which von Wright warns. The general mood today praises and privileges scientific knowledge. We all know, for example, that there are psychological differences between men and women. Then suddenly some scientific research reveals that there are differences between the brains of the two sexes. Now we know the previous fact scientifically – as if we did not really know it before. When something is scientifically proven or a result of scientific research, the hope that everything can become knowable in this way arises. When this happens it should be remembered, as von Wright points out, that knowledge without understanding cannot cope with all the problems that are important for us, and may in fact lead to darkness.

The methodological monism of reductive or eliminative naturalism is opposed to the kind of view which the classical naturalists, in their refusal to eliminate the seeming plurality of the world, strove for. For a classical naturalist, and for a modern naturalist like Rorty, there are no fundamental distinctions like science / non-science or natural / non-natural. A classical naturalist would say that whatever is, in whatever way that it is, belongs to nature.\textsuperscript{46} Von Wright’s description of our times, in which understanding has suffered a loss in favor of knowledge, is an example of a situation in which overemphasizing one method (which is taken to be exemplary and applicable to every aspect of nature) leads to the ignorance of something which is real and important for human life as well. An eliminativist attitude towards mental phenomena is a concrete example of this. It could be claimed that the threat which science, with its promise of factual knowledge, poses for philosophy is a result of a general symptomology of the intellectual mood of our times.

\textsuperscript{45} Von Wright, 2000b, 280, my emphasis.

\textsuperscript{46} An interesting example of modern “liberal naturalism” can be found in Buchler, 1966.
Philosophers who want to turn philosophy of mind into a scientific enterprise claim that non-scientific discourses, philosophy among them, are mere loose talk and they are not telling the truth about the world. In claiming this, they ignore the fact that non-scientific discourses have functions and serve purposes which are important for the cognitive lives of humans, whether or not these discourses count as knowledge revealing the “ultimate truth” about reality. Labeling these non-scientific discourses as “loose talk” because they are not telling the “truth” about the world is to seriously overestimate the importance of the role of science for human life. To label non-scientific discourses as “loose talk” is to downgrade the fact that human life is based on these discourses in a very concrete sense. The scientific method, while certainly successful, beneficial, and important with respect to many problems, does not help in answering what people should do with their lives or with other “deep” questions. Science has its proper place, but it should not infiltrate areas which are clearly outside its domain. It is doubtful that the scientific method could offer the kind of overall picture of the human condition which many desire. Since one purpose of philosophy can be taken to be the evaluation of the human condition in its entirety, i.e. to try to formulate a coherent general picture, a scientific philosophy which simply ignores or eliminates aspects of the human condition and dictates that questions have to be answered in a certain way by following specific methods falls short of being good philosophy and becomes dogmatism.

Whereas a scientific conception of philosophy – reductive naturalism – has a firm place in the modern philosophy of mind, the demand for plurality is expressed by many contemporary philosophers – people as diverse as John Dupre, Barry Stroud, Jennifer Hornsby, Hilary Putnam, Noam Chomsky, Stephen Stich and John Mcdowell, to name just a few. Some of their views resemble the views of the classical naturalists and emphasize the need for intellectual and methodological diversity in the study of human reality. Davidson and

47 Mcdowell (2004, 95) defines his view by saying that: “[…]) liberal naturalism – does not accept that to reveal thinking and knowing as natural, we need to integrate into the realm of law the frame within which the concepts of thinking and knowing function. All we need is to stress that they are concepts of occurrences and states in our lives.” This version of naturalism should be contrasted with restrictive naturalism, which “[…] aims to naturalize the concepts of thinking and knowing by forcing the conceptual structure in which they belong into the framework of the realm of law.” (ibid.) As we have seen, this latter view is something which, for example, Fodor defends. Putnam (1999, 38) writes that: “The metaphysical realignment I propose involves acquiescence in a plurality of conceptual resources, of different and not mutually reducible vocabularies… coupled with a return not to dualism but to the “natural realism of the common man.” According to Stich (1996, 197): “On the open ended pluralistic picture I would urge, we should not expect that we could find the sort of criterion that the naturalist seeks… the ontology embraced by the best of our physical and biological sciences… is astoundingly diverse… if we throw in the ontology of the social sciences, the list seems even more heterogeneous… according to the naturalist all of these (or at least all that are really respectable) must stand in some special relation to the properties of physics. This strikes me as singularly implausible proposal.” Finally, Chomsky (2000, 77) notes that: “…a naturalistic approach does not exclude other ways of trying to comprehend the world.” This remark is often neglected by those who claim that science is the highest path to truth.
von Wright are clear examples of thinkers who defend the need to pursue methodological diversity. Davidson, for example, accepts the Rortyan claim according to which we are equally in touch with reality through different ways of describing it.\textsuperscript{48} As long as the debate about the demands of naturalism remains unresolved, the question of what counts as a sufficiently naturalistic philosophy of mind is also open. If all philosophical problems of the mind can be solved by discovering more facts, the success of science is the reason why a scientific philosophy of mind threatens to replace “pure” philosophy of mind. If all questions about the mind are questions which can be solved by doing empirical research, then the unquestionable success of, for example, neuroscience is one reason why it could replace pure philosophy of mind – conceptual analysis – entirely. If all problems of the mind are such that the best way to study them is in terms of natural science, then the discipline called philosophy of mind can and perhaps should be forgotten.\textsuperscript{49} If mental phenomena require scientific vindication, then the threat of intentional irrealism exists insofar as mental phenomena cannot be vindicated in a scientifically respectable way.

In evaluating the importance of classical naturalistic philosophers like Davidson or von Wright, it is crucially important to be clear on the question of what currently counts as legitimate philosophy. My claim is that their views are criticized mostly because they do not fit the prevalent naturalistic consensus. Their conception of philosophy is in conflict with the modern view concerning what philosophy should be. But if the principles that make up the prevailing naturalistic consensus are a mistake, then a major part of the criticism against Davidson and von Wright is ill-founded as well.

1.1.1 Science gone too far – eliminative materialism

In their denial of the “supernatural”, Davidson and von Wright subscribe to a form of non-reductive naturalism. Other forms of naturalism deny much more than the existence of supernatural entities and, allegedly, are threatening to eliminate all that is human.\textsuperscript{50} Eliminative materialism draws a dramatic conclusion: mental phenomena cannot be reduced to the phenomena which “hard” sciences talk about; therefore mental phenomena do not exist.

\textsuperscript{48} Davidson, 1999.
\textsuperscript{49} As Kim and others have noted, the role of philosophy is changing. This creates the question of what the current and future task of philosophers is. It also raises a question concerning the relevance of past philosophers like Davidson and von Wright. The question of the future of philosophy is attracting the attention of philosophers, as is shown by the titles of the books like The Future of Philosophy: Towards the Twenty-First Century (Oliver, 1998) or The Future For Philosophy (Leiter, 2004). Philosophers who have written on the subject include Searle, 1999 and Kim, 2004.
\textsuperscript{50} See Rudder-Baker, 1987.
The main reason why mental phenomena may have to be eliminated from the one “true ontology” is that they do not fit well into the picture created by science. The progress of science is a threat to pure philosophy, but threatens also phenomena which do not keep up with this progress. Paul Churchland writes:

FP [folk-psychology] shows no sign of being smoothly integrable with the emerging synthesis of the several physical, chemical, biological, physiological and neurocomputational sciences. Since active coherence with the rest of what we presume to know is a central measure of credibility for any theory, FP’s emerging wallflower status bodes ill for its future.⁵¹

What Churchland is suggesting is not that active coherence with common-sense knowledge is a measure of credibility, because on that criterion the status of folk-psychology would be secure. What we really know is determined by science, and active coherence or non-coherence with scientific knowledge is the final measure of credibility. Eliminative materialism reveals the consequences of a view privileging the results of science.⁵² Even deep intuitions about ourselves should be ignored if they cannot be confirmed by science or made to fit with scientific knowledge. This rejection can be sustained, but not without doing great injustice to views which undeniably have an enormous role in human life. If philosophical honesty is respected and eliminativism is taken seriously, it should have consequences also outside the circles of dry academic philosophy.⁵³

Naturalists who try to avoid these eliminative conclusions have proposed theories about the mind which are a mixture of scientific results and pure philosophical speculation. The reductive theories of Fodor or Dretske and eliminative theories of the Churchlands or Stich are themselves all highly speculative and not views confirmed by empirical evidence.⁵⁴ There seems to be a great deal of confusion in the contemporary philosophy of mind, all of which comes back to this: philosophers are not clear about the nature of their philosophical investigations. Therefore they try to steer a middle course between “pure conceptual analysis” and “pure empirical philosophy”. The result is a position which can be labeled as empirical

⁵¹ Churchland, 1998a, 8. See also Churchland, 1983.
⁵² Putnam (1988, 59) criticizes the Churchlands on the grounds that “Their whole argument turns on the following inference: if the instances of X do not have something in common which is scientifically describable... then X is a "mythological entity."
⁵³ For a discussion of philosophical honesty, see Hertzberg, 2006.
⁵⁴ Although Churchlands emphasize that most philosophers have been thinking too much and have not done enough experiments, their own speculations, which sometimes reach the level of science fiction, fit poorly the conception of philosophy which they seem to defend. At the same time, it needs to be acknowledged that they intend to base their views on empirical findings and have certainly produced interesting research results on which their more philosophical views rely. Whether the research results really imply everything that the Churchlands claim is questionable, as it often is with scientific findings that have competitors.
The proposals of reductive and eliminative naturalists transcend what is verifiable and postulate phenomena like ‘the language of thought’, which are metaphysical constructions meant to vindicate what already lies before everyone’s eyes. These proposals or constructions are optimistic assumptions or predictions that are put forward in a style that gives the impression that the views, because they are “scientific”, amount to established facts. Philosophy of mind has become a battleground where pseudo-empirical views, hybrids of science and philosophy, compete. Paul Churchland ends his book *The Engine of Reason the Seat of the Soul* by claiming:

You came to this book assuming that the basic units of human cognition are states such as thoughts, beliefs, perceptions, desires... these bedrock assumptions are probably mistaken. In humans, and in animals generally, it is now modestly plain that the basic unit of cognition is the *activation vector*. It is now fairly clear that the basic unit of computation is the *vector-to-vector-transformation*.

The ‘activation vector’ as a basic unit of cognition is an example of what we scientifically claim to know, and our folk-psychological conception of ourselves must be made to fit with this knowledge or be dismissed as another myth. We are told that it is now modestly plain that the “basic unit of cognition” has been found and therefore: “...recent science already suggests that... folk-psychology... fails... to capture the basic kinematics and dynamics of human... cognition.”

The conclusion is:

[...] our commonsense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced... by completed neuroscience.

This is a radical claim. Yet, we also hear: “We are still too ignorant to insist that hypotheses [about activation vectors] will prove adequate to explain all of the representational capacities of mind. But neither can we insist that they are doomed to prove inadequate.” Which view is

---

55 I borrow the term from Glock, 2006.
56 Churchland, 1995, 322. The view that folk-psychological conceptions should be replaced with a neuroscientific theory because the truth of the latter is already well confirmed is repeated in various places. See Churchland, 1988, 294, 351, Churchland and Churchland, 1998, 10, 13, 41, 257-258, 272-273. Whereas eliminative materialists claim that the basic unit of human cognition is vector-to-vector-transformation, Fodor (1998), who holds a very different view about the nature of human cognition, claims that “nothing is true” in Paul Churchland’s book mentioned above. This shows that although scientifically minded philosophers claim that they have well-confirmed empirical theories about the mind, the acceptance-rate of those theories among other, even like-minded philosophers is, to say the least, quite modest.
57 Churchland, 1998c, 10.
58 Churchland, 1981, 67, my emphasis.
59 Churchland and Churchland, 1998a, 76. The difference in tone is remarkable. Sometimes, for example in Churchland, 1998c, it is argued that it is already a scientifically proven fact that folk-psychology is a false theory, which is just about to be replaced by a neuroscientific theory. Sometimes, like in the quote given here, it is only claimed that we cannot at the moment prove that the new theory won’t come out as an inadequate one.
correct? Because the claims of eliminative materialism often involve predictions about the future, it is difficult or impossible to evaluate their correctness. My claim is that one way to measure their correctness is to consider how well the claims fit with what we currently know about ourselves. As philosophers we should not be mesmerized by science and should resist the temptation to put the scientific cart before the horse of the ontology of actual human life.

We have reasons to believe that our knowledge about the brain increases our understanding of the nature and functioning of the mind as well, to some extent. But this modest view is something entirely different from a philosophical or ideological view according to which all descriptive vocabularies must be made to fit with the ontology of physicalism through reduction. This latter claim is implicit in naturalistic philosophy, be it a reductive version like Fodor’s or eliminative version like Churchlands’, and in its explicit claim that there is no room for pure philosophy any longer. Of course the results of neuroscience cannot and should not be dismissed. Notable progress has been made in a relatively short period of time. This progress is bound to have both negative and positive effects for philosophy. What is worrying, as von Wright has noted, is that human understanding may suffer a great loss if the importance of scientific knowledge is over-emphasized. As a consequence of this, scientists may draw the wrong conclusions from their evidence and overestimate the importance of their scientific findings. A scientistic atmosphere in society encourages this kind of activity and scientists as well as many philosophers are not immune to societal influences. A possible consequence is that scientific results and their implications infiltrate areas in which scientific solutions are, or so many may feel at least, inappropriate and even harmful. Scientists and scientifically inspired philosophers may ignore philosophical problems by altogether dismissing them or by transforming these problems into scientific questions.

Although reductive and eliminative naturalists often claim that the purpose of science is to provide a superior conception of the world, and that therefore the results of science are privileged, “pragmatic considerations” also influence these forms of naturalism to an increasing extent.⁶⁰ In the philosophy of mind, pragmatism is often associated with some form

---

Because the views are roughly from the same time-period, the chronological order of the suggestions does not explain this change of mind.

⁶⁰ “Pragmatism” is obviously such a complicated position that it would not be beneficial to try to define its nature here in detail. The crucial distinction between naturalism and pragmatism that I want to make is merely that whereas science is the highest path to truth according to the former position, the latter position, at least in its Rortyan form, suggests that science is not privileged because of metaphysical reasons. It does not better capture the ultimate truth about reality and does not bring us closer to reality than other descriptive systems. Insofar as contemporary naturalism is really moving towards pragmatism, it needs to get rid of the idea that science
of non-reductive naturalism, and it is unclear how reductive or eliminative naturalism can accept pragmatist conclusions without rejecting the views about the importance and nature of reduction which are essential to these positions. In Paul Churchland’s case, the position of eliminative materialism and the conception of reduction on which the position is based sometimes allow room for a form of pragmatism. The reason for this is the observation that if a reducing theory always eliminates the theory which is reduced, the consequence is that all scientific theories have been falsified and that all future theories will, most likely, also turn out to be false. If old scientific theories are eliminated in favor of new ones, at any given time we face an unattractive state of skepticism because all current theories are possible candidates for falsification and thus for elimination by future theories, just as much as the old theories were. As Churchland notes:

So many past theories, rightly judged excellent in their time, have since proved to be false. And their current successors, though even better founded, seem but the next step in a probably endless and not obviously convergent journey.61

To insist or suppose that now we have arrived at true theories about reality lacks a rationale. True, scientific theories must be largely correct if the ability to predict, control or manipulate the course of events is the measure of truth. But doesn’t this mean that the correctness of these theories relates to their usefulness with respect to certain kinds of tasks? Yet, there are aspects of reality – mental reality is an obvious and perhaps the best example – for which we do not have theories of strict manipulation, control or prediction. Those who are enthusiastic about scientific psychology perhaps say that psychological theories allow for the prediction and explanation of human behavior. But, at least at present, the explanation or prediction of the behavior of an individual is impossible; this should be indisputable. Prediction and explanation of human behavior is of a totally different kind from prediction and explanation in the natural sciences. The answer to the question “Why was it that a certain individual did x at time t?” is always speculative, unlike explanations in natural science.

The claim that theories which depict us as we “truly are” are forthcoming lacks a foundation.62 As Chomsky among others has argued, it is possible that the cognitive capacities of humans are simply too modest to understand certain aspects of reality or to produce a “final theory of everything”. The nature of reality may just be too complex for

---

61 Churchland, 1989, 140.
62 The expression “as we truly are” is from Churchland, 1995, 324.
human capacities. This may be so especially when we are trying to understand ourselves. Perhaps our minds could not really comprehend a situation in which their workings could be strictly predicted, controlled or manipulated. This kind of conclusion would then not be a **cognitively live option** for humans. It is perfectly possible that the problem of consciousness and the problem of mental causation will remain mysteries which will never be scientifically solved. Some philosophers have already reached this conclusion.

To avoid the skeptical dead end, Churchland concludes that we should: “…remove the goal itself – a unique truth – as well as any sure means of getting there” and that “…the proper course to pursue in epistemology lies in the direction of a highly naturalistic and pluralistic form of **pragmatism**.” We are told: “Service to our practical purposes is the only justification that will really count in the end.” This sounds like the statement of a pragmatist. It is something that Putnam and Rorty, among others, have been stressing for a long time. Whereas eliminative materialists criticize folk-psychology on the grounds that this conceptual framework does not describe us as we truly are, it is also acknowledged that folk-psychology nevertheless has “considerable predictive success”. Predictive success is useful and given this service to practical purposes, on Churchland’s own account this should, pace the eliminative polemics, justify the retention of folk-psychological notions. If one accepts, as Churchland does, that: a) all knowledge is speculative and revisable; and b) “…presumptive knowledge earns that status by allowing us to anticipate, to explain, and in general to navigate and to manipulate phenomena within the domain thereby grasped, whether natural or social”, then how can the claim of folk-psychology’s radical falsity be supported? If it is not the reducibility to hard sciences which determines the fate of folk-psychology but the service to our practical purposes, then how could folk-psychology be a candidate for elimination? We can try to anticipate and explain our own behavior or that of others in terms of activation.

---

63 Famous quantum physicist Anton Zeilinger has recently emphasized that he cannot really understand any longer the results of quantum physics, or what the nature of reality is, if the results are followed to their conclusion. Different kinds of opinions obviously exist, as the title of physicist Steven Weinberg’s book *Dreams of a Final Theory* suggests.

64 By considering the nature of certain mental illnesses like schizophrenia, where people sometimes think that somebody else is controlling them or giving them orders, it is perhaps possible to get a feeble glimpse of the problematic situation in which human mind could be if somebody could actually strictly manipulate its autonomy. This obviously is not an argument against the possibility of that kind of manipulation, but the question is whether the resulting situation would be comprehensible for human cognition.

65 See for example, Mcginn, 1999.

66 Churchland, 1989, 194, my emphasis.

67 Churchland, 1995, 324.

68 Churchland keeps also emphasizing that there is no final theory and no final science, and that we should let go of the utopian attitude when it comes to our long-term cognitive progress. See for example Churchland, 1998a, 43-44 and Churchland, 1989, 151.

69 Churchland, 1988, 395.

vectors. Then we can attempt to anticipate and explain the same phenomena in folk-
psychological terms. It does not need to be emphasized which mode of explanation succeeds.

If service to practical purposes is what counts, what does it mean to claim that: “The
function of science... is to provide us with a superior... conception of the world”? In what
sense is this conception superior? In the sense that science is the cognitive activity best
serving our practical purposes? As far as I understand, this is not the sense which Churchland
intends. He suggests that science provides a superior conception of the world because it
describes the true nature of reality. But is it possible to maintain the distinction between
“reality as it is” and “reality as it seems to us” in the context of pragmatism? This distinction
does not make good sense, and this is perhaps reason why Churchland describes himself as
“at least a closet pragmatist”.

What the consequences of pragmatism for folk-psychology are is an interesting question, because if service to practical purposes is what counts then status of
t folk-psychology or “common-sense” should not be under threat; on the contrary. This would
mean that the threat from science to mental phenomena and common-sense is over-
emphasized. The serious threat to philosophy is also averted if the goal of unique truth and the
putative sure means proposed for getting there are rejected. What is important is not the truth
of a theory, but how useful a given theory is. According to the pragmatist criterion, the
“goodness” of a theory depends on its practical usefulness. If this criterion is accepted, then
many philosophical theories about the mind can compete with scientific theories.

Whereas both Paul and Patricia Churchland on occasions worry that: “...folk-
psychology constitute[s] a radically false account of the cognitive activity of humans...” or
claim that “...FP is approaching the brink of falsification...” or insist that folk-psychology is
a boat “...leaking at every seam”, they nevertheless do not want to rule out a “pragmatic
toleration” of a strictly false, yet a highly useful folk-psychological causal explanatory
theory. It is admitted that folk-psychology will be a crucial element in a neuroscientific
theory of the future and that: “Despite the occasional polemics...the primary lesson is not that
FP is already doomed, or that our current social practices are about to be swept away.” On
Churchland’s pragmatist view, there should be no other false / true distinction than the
pragmatic criterion – according to which service to practical purposes is what counts. The

---

71 Churchland, 1979, 2.
72 Churchland, 2001a, 103.
73 Churchland, 1998
74 Churchland, 1998b, 33.
75 Churchland, 1998b, 36.
76 For this view see McCauley, 1996.
77 Churchland, 1998b, 38.
importance of folk-psychology and the need to maintain this framework should be evident if the pragmatic criterion is what counts. Reductive or eliminative naturalists seem to insist that god’s point of view remains beyond human reach because scientific knowledge is always fallible and open to revision, and yet they want to claim that science provides a superior conception of reality.

The distinction between a true, scientific description of human mentality and a description which is useful but false remains utterly unclear if service to practical purposes is what counts in the end. It remains a mystery how an expression like “pragmatic toleration” of a theory should be understood, and the status of a tolerant view is unclear if we accept the view that a scientific “utopian theory” is always privileged. The pragmatic toleration of a false theory resembles Quine’s double standard with respect to physical and mental notions. But it should be asked, as Davidson once did: “What is the point of saying that intentional idioms are ‘second class’ if we are going to go on using them?” Perhaps it could be claimed that all naturalists are willing to admit this: there are important “ways of speaking”, and we cannot get rid of them because they are so useful. But naturalists would also claim that these are not scientific views, and ultimately they are literally just ways of speaking. We could ask: So what? What is the point of emphasizing that everyone agrees that certain vocabularies are useful and indispensable, but at the same time according them a second-grade status?

What can be said about the contemporary challenge for the philosophy of mind is this: Whereas eliminative materialists like the earlier Quine, Rorty, Stich, as well as the Churchlands have now reached the conclusion that common-sense mental notions are too useful to be dispensed with and therefore the proper course of epistemology lies in the direction of a kind of pragmatism, their desire to turn philosophy into a scientific enterprise or altogether dispense with philosophy in favor of “hard science” is nevertheless evident. That naturalism is, in some sense, going through a “pragmatist turn” is interesting, but the real consequences of this can be evaluated only in the future. Here I have very briefly described

78 A pluralistic form of naturalistic pragmatism gains ground in Churchland’s views from time to time, but the grip of sacred science is also firm. Paul Churchland (1989, 52, 1988, 47) claims that, when it comes to the human cognitive life, there is a “...correct account that a utopian theory will eventually provide” and that the “…a priori probability of eliminative materialism is not lower, but substantially higher than any of its competitors.” The latter view is defended with an obscure claim. It is claimed that there are more ways of giving a successful neuroscientific account which does not mirror the ontology of folk-psychology than giving a neuroscientific account which does mirror this ontology; thus it is more likely that eliminative materialism is true than that it is not. This is of course pure speculation.

79 As quoted by Putnam, 1987, 70.

80 For Quine’s eliminative views see Quine, 1975, 1985. For a change of mind, see, Quine 1990 and 1995a. For Rorty’s eliminative views, see Rorty, 1965, 1970, 1970a, 1977 where he argues that scientific progress will rid us of the need to establish identities between brain states and mental states. For a change of mind see Rorty, 1999. For Stich’s eliminativism see, Stich, 1983. For a change of mind see Stich, 1996.
how the position of eliminative materialism, the most scientifically oriented philosophy of mind, faces problems if its main message is taken seriously. There seems to be an intrinsic conflict in the views of eliminative and reductive naturalists, a conflict which is summed up by Davidson’s question about what the real point of the scientific / non-scientific distinction is. Despite this conflict, the challenge of naturalism is a problem for any position, such as Davidson’s and von Wright’s, which is highly non-empirical and yet tries to seriously challenge views based on empirical evidence. The relationship between these two understandings of philosophy is one of the most pressing (although largely ignored) questions of modern metaphilosophy.

1.2 A Davidsonian – von Wrightian response to the challenge

The philosophies of Davidson and von Wright are rarely compared to each other. From the textual evidence it can be seen that these philosophers did not challenge each other’s views directly in print. The lack of discussion between these two philosophers and lack of comparison of their views by scholars are surprising given that their philosophical careers spanned almost the same period of time and touched on many of the same philosophical issues. One important exception, which has also created discussion, is their work in the philosophy of action where, so the standard interpretation goes, Davidson and von Wright held opposite views during the 1960s and 1970s. This interpretation is still held by many commentators. It is noted for example that: “[…] the only real resistance to Davidson’s ruling influence over the Philosophy of Action has come from various Wittgensteinian camps, most prominently that of G.H. von Wright and his pupil Frederick Stoutland”.81 Views like this show that on those rare occasions when Davidson and von Wright have been mentioned together, the focus has been on the differences of their views. A common interpretation is that, at least in the philosophy of action, a “Davidsonian” causal approach must be separated from a “Wittgensteinian” non-causal approach. The differences and similarities of Davidson and von Wright’s views will be discussed later in detail. Here I focus on the question of what, according to these philosophers, the purpose and nature of philosophy is. With respect to this question, Davidson and von Wright do think alike.

Davidson’s and von Wright’s positions are vulnerable to the challenge that science poses for the philosophy of mind. My claim is that their views in the philosophy of mind are

criticized and largely ignored because their overall conception of what philosophy is does not meet the contemporary requirements of what philosophy should be. Davidson’s philosophy of mind was widely discussed earlier, but now his views are rejected to the extent that there are basically no philosophers who would today accept, far less defend, his theory of mind. Von Wright’s case is worse. There is literally no discussion about his views concerning philosophy of mind. It is surprising that even the few von Wright scholars in Finland have almost completely ignored this aspect of his philosophy.

Because the reason that the views of Davidson and von Wright are not well accepted among modern philosophers of mind traces back to their understanding about the nature of philosophy, it is crucially important to be clear on their views on this issue. I suggest that the contemporary challenge for philosophy of mind can be overcome by considering what one specific purpose of philosophy is, or could be, and by showing that a scientific philosophy does not and cannot meet this purpose. I do not want to claim that von Wright or Davidson are trying to meet the naturalistic challenge on purpose although they – especially von Wright – do express severe criticisms of the scientistic tendency in modern philosophy. What I am claiming instead is that it is possible to take cues from their views in showing the challenge can be overcome if it is taken up.

It is fair to say that Von Wright’s and Davidson’s views about the mind / brain are highly non-empirical. These views are very removed, both in their style and in their goals, from the kind of philosophy that works like Patricia Churchland’s Neurophilosophy, Paul Churchland’s Neurophilosophy at Work or John Bickle’s Philosophy and Neuroscience: A Ruthlessly Reductive Account represent. The number of philosophical books which are strongly based on empirical evidence is increasing. As works of philosophy, they are quite different from many of the philosophical classics of the 20th century. This is not a point that needs further emphasis; it merely shows that the nature of philosophy is changing. Perhaps this observation supports von Wright’s view that in modern philosophy, Russell’s view about the nature of philosophy has triumphed.

From the perspective of naturalism, the speculative character and non-empirical nature of Davidson’s and von Wright’s views makes them straightforwardly false, or at best uninteresting or irrelevant. As the titles of the Churchlands’ or Bickle’s books show, the

---

82 One notable exception is Campbell, 2008.
83 The most up-to-date discussion of Von Wright’s philosophy is Stoutland, 2009, which however does not really discuss many aspects of von Wright’s philosophy of mind.
84 It should be obvious that I am not saying that the challenge can be dismissed. I am merely showing one way how to think about the challenge and its relation to philosophy.
integration of philosophical views of the mind with scientific work, especially with neuroscientific results, is one way to do philosophy. Philosophical claims are, to an increasing extent, backed up by scientific evidence and, so it is argued, cannot stand on firm ground without such evidence. This predicts a dim future for the views of Davidson and von Wright. Their relevance for the modern discussion is judged as historical, at best. This kind of view reflects the opinion of the majority of scholars studying their work, and generally describes the different orientations of contemporary philosophers of mind. It is argued that philosophers of mind should simply be interested in different questions than those Davidson and von Wright considered, and should seriously focus on empirical results in the future.

Although von Wright and Davidson do not take empirical results into consideration to any great degree, their views are not openly antiscientific. They belong to the naturalist consensus in the sense that they reject ontological dualism and they do not put forward claims which would be in direct conflict with scientific knowledge. This being said, the positions of these two philosophers could be still labeled as antiscientific. This could be done because they do not take into account well enough the results that science has provided, and because they criticize the idea of the unity of science and are somewhat critical of the idea of scientific philosophy. Scientifically oriented philosophers could argue that some of the philosophical proposals of von Wright and Davidson are simply not true, and that these views are antiscientific and false precisely because they are at odds with what we already know about the mind due to science. As will become clear, with respect to certain suggestions of von Wright and Davidson this complaint seems valid. But the reason that the views of these two philosophers are nevertheless important and relevant for the current debates about the mind is that they are doing philosophy of mind instead of neuroscience, cognitive science, or any kind of “philosophy” which tries to mimic these and which is often guided by the abstract demands of scientistic ideology.

A critic could claim that philosophy of mind is an interesting approach only if philosophy can make a contribution to the problems of mind which differs from the contributions made by science. On the other hand, a defender of philosophy could claim that the importance of philosophy is based on the fact, or perhaps more modestly on the view, that philosophy is a different kind of intellectual activity than science. Naturalists would perhaps grant this; yet, as we have seen, they would claim that science, as a human activity, stands on a higher level than philosophy. We can recall Patricia Churchland’s view that philosophers

85 There are modern philosophers who defend Cartesian dualism. See e.g. Foster, 1991. On one interpretation these philosophers could be rightly labeled as antinaturalists.
should free themselves from the “narrow canyons” of the commonsense conception of the world. Commonsense conceptions and the kind of philosophy which relies on them remain in the narrow canyons whereas science, allegedly, pushes back the “bounds of darkness”. Whether philosophy of mind can be seen as an autonomous field and whether the views of von Wright and Davidson are relevant depends on how one understands the relation between philosophy and science. It is again the nature of philosophical method which is at stake.

Although the positions of von Wright and Davidson are vulnerable to the challenge made by science, my claim is that they overcome – or more importantly could overcome – this challenge because of their similar views on what the purpose of philosophy is. Those who have a differing view about the nature or purpose of philosophy may find most of the claims of these two philosophers uninteresting or irrelevant for the contemporary “philosophy” of mind. As was described in section 1.1, differences of opinion about the purpose of philosophy divide contemporary philosophers of mind. The existence of these differences is something which should be recognized when comparing philosophical accounts of the mind against each other and when evaluating their merits or correctness. These differences of opinion concretely influence the way in which current philosophy of mind is done, in what direction philosophy of mind should allegedly proceed, and also what kind of solutions are currently favored. But surely the possible merits of a philosophical position should be considered without having to consider whether they meet certain abstract standards which happen to be currently favored. The standards are result of the mood of our times and therefore depend on sociological, political, economical and other factors which are arbitrary from a philosophical point of view.

The different views about the purpose of philosophy have to be acknowledged and accepted, because it is an absurd suggestion that philosophizing as a human activity has only one purpose, that there is an “essence of philosophy” as an activity. It is impossible to settle the question of what the purpose of philosophy is or what the purpose of philosophy should be. Because philosophy has no clearly defined nature, one could say that its purpose or nature is a matter of personal opinion or intuition. Instead of trying to convince the reader that the Davidsonian – von Wrightian understanding of the purpose of philosophy is the right one, I merely describe their understanding in order to justify and clarify the consequent approach which they have adopted to the problems about the mind.\textsuperscript{86} This description makes it easier to see where the real conflict between a naturalistic philosophy and a Davidsonian–von Wrightian approach lies. My purpose is not to consider what Davidson’s and von Wright’s

\textsuperscript{86} What would it even mean to say that this is the right approach, that this is what one should think about philosophy?
rationales for their conception of philosophy are. The question about what kind of conception of philosophy one should hold is a question to be judged ultimately by the reader as she sees fit. There is no easy answer to the question of why a certain philosopher finds a specific conception of philosophy congenial, not to mention why a certain conception should be found congenial.

I want to suggest that there is a philosophical figure that is an important inspiration for both Davidson and von Wright and, more importantly, whose understanding of the nature and purpose of philosophy clarifies the respective views of these two philosophers. This figure is Wittgenstein. The influence of Wittgenstein on von Wright is evident. Von Wright was Wittgenstein’s student, friend, follower in Cambridge and, according to his own words, as Wittgensteinian as possible without being a Wittgensteinian. Von Wright also noted that he “learned philosophy from Wittgenstein” and that he learned more from him as a moral example than from anybody else. Despite all this, not much attention has been paid to the fact that understanding von Wright’s Wittgensteinian conception of philosophy is crucially important for understanding his work in general and his work in the philosophy of mind in particular. Wittgenstein’s influence for the development of von Wright’s views is a theme which, although studied, has not been explored enough. This being said, von Wright claimed that his philosophical work, his way of stating philosophical problems and arriving at conclusions, was very different from Wittgenstein. It is true that von Wright’s work in logic is alien to the “Wittgensteinian spirit”; I think, however, that von Wright’s later work in the philosophy of mind comes closer to this spirit. It is clear that von Wright shares Wittgenstein’s view about the nature of philosophy to a great extent, although their philosophical styles may differ.

Davidson, on the other hand, has been credited or discredited in light of the fact that his views on action and its explanation revived the interest in causal theories, and that on these questions Davidson’s views are opposite to those of Wittgenstein. According to

---

87 Finding this out would be a very interesting task, but it would take the discussion too far, towards issues relating to Davidson’s and von Wright’s life stories. There are also people who can answer this question without the kind of speculation that I would offer. This being said, in the appendix I do want to offer a brief suggestion about how von Wright’s philosophy of mind could be related to his “social humanism”. Stoutland (2006b) has noted, correctly in my opinion, that In the Shadow of Descartes is one of the works resulting from von Wright’s general shift of attitude. But he has not explored the question how von Wright’s social humanism is related to his philosophy of mind.

88 At the same time von Wright (2001, 179) notes that: “[...] neither my style of writing nor my treatment of particular problems have much in common with Wittgenstein’s.” However, I believe that if one considers the style and the arguments of In the Shadow of Descartes, it is easy to see that this self-observation of von Wright is not entirely correct.
Stoutland, Davidson is near the top of the enemies list of the Wittgensteinians. Stoutland is, however, one of the few commentators who see important connections between Davidson and Wittgenstein. In this sense, he is also one of the most perspicacious interpreters of Davidson, capturing the spirit of Davidson’s work – and his analysis certainly casts doubt on the conventional wisdom that Davidson’s causal theory of action stands in sharp contrast to non-causalist accounts. I shall not here repeat Stoutland’s arguments, which I find very convincing. Against Stoutland’s interpretation, P. M. S. Hacker, the famous Wittgenstein scholar, writes that Davidson’s “[…] constructive theory, anomalous monism, won many adherents and was patently at odds with Wittgenstein’s conception of action and its explanation.” When the details of AM are discussed I will show that Hacker’s negative view of the status of anomalous monism is seriously questionable. The critical view of Hacker is no surprise, given the fact that he is among the “Wittgensteinians” to whom Stoutland refers. I think, however, that Hacker is misinterpreting Davidson’s views and that these views actually come close to a conception of philosophy which Hacker, as a Wittgensteinian, supports. Certain similarities between Davidson and Wittgenstein, which will be presented here, have gone unnoticed by the commentators. My purpose is to show that Davidson’s philosophy of mind is clearly influenced by Wittgensteinian insights. It is interesting to notice that certain points made by Wittgenstein are repeated almost literally by Davidson.

My interpretation of von Wright and Davidson as philosophers continuing the work of Wittgenstein in the philosophy of mind places them in the center of a recent discussion. This debate is interesting, as it shows a return to the very problems which Wittgenstein emphasized in his times and, so I believe, shows the relevance of a Wittgensteinian conception also for contemporary discussion. There are reasons to think that the same problems to which Wittgenstein drew our attention survive in a modified form in the views of many modern philosophers and scientists studying the mind, brain, or mind-brain. This would mean that the importance of Wittgensteinian insights may not yet have reached their peak. The interesting discussion to which I am referring was started by the publication of Philosophical

---

89 Stoutland, 2008.  
90 Stoutland was one of the earliest commentators to suggest that Davidson’s position may have epiphenomenal consequences, and earlier he argued that Davidson’s views are in clear conflict with both Wittgenstein’s and von Wright’s views. See Stoutland, 1986. For a change of mind with respect to Davidson and von Wright, see Stoutland, 1999.  
91 For a similar view about Davidson’s position see also Mcintyre, 2004.  
93 At times Davidson acknowledges Wittgenstein’s influence. Stoutland (2008) tells of how he once asked Davidson whether he should be seen as a “closet Wittgensteinian”, to which Davidson replied “I don’t know about the closet.”
Foundations of Neuroscience by Hacker and Bennett. In this book, which von Wright incidentally described as something which “[…] will certainly, for a long time to come, be the most important contribution to the mind–body problem which there is”, Bennett and Hacker interpret the recent results of neuroscience and neurophilosophy from the perspective of Wittgenstein’s later philosophy. Their main claim is that modern neuroscience is guilty of serious conceptual confusions and, as a metaphilosophical suggestion, they claim that one important task of a philosopher is to draw attention to such confusions. Two philosophers who were especially targeted in the book were Dennett and Searle. This in turn led to an intense debate resulting in the publication of Neuroscience and Philosophy, in which the replies of Dennett and Searle are presented.

The debate is fascinating, since it shows what happens when philosophers with very different presuppositions and different intellectual agendas try to find common ground. What the debate shows is that, at least in some cases, finding such ground is simply not possible. Philosophers who disagree about the nature and purpose of philosophy find it difficult to find a common ground also on the more substantial philosophical topics. It is telling that whereas von Wright held a high opinion of the Philosophical Foundations of Neuroscience, calling it one of the most important contributions to the mind–body problem, Dennett’s view about the same book is: “I found nothing new in their book.” Because of the foundational nature of the disagreement, the debate will certainly continue between philosophers like Bennett and Hacker on the one side and philosophers like Dennett, Searle and the Churchlands on the other. Sometimes a common enemy, in this case Bennett and Hacker, provides a reason for philosophers disagreeing with each other to join forces. Searle and Dennett have a serious dispute with each other and Searle’s views are very different from Paul Churchland’s position as well. Yet, all vigorously want to oppose the “Wittgensteinians”. Sometimes it seems that the only thing which brings reductive naturalists together is their hostility towards Wittgensteinian insights.

What is interesting about this debate from the viewpoint of this work is the following: In his reply to Bennett and Hacker, Searle notes: “The best way to understand [Bennett’s and Hacker’s] book is to see it as an application of Wittgenstein’s philosophy of mind to

---

95 Von Wright’s appraisal can be found in the inside cover of the book.
96 Bennett (et al.), 2007.
97 Dennett, 2007, 75 fn. 4. For this comment Dennett may have personal reasons, because he considers that Philosophical Foundations of Neuroscience contains a “remarkably insulting attack” against him. I am tempted to ask: Can philosophical disputes be settled objectively in this kind of setting?
98 Paul Churchland (2005a) has critically commented on the conclusions of the original book.
contemporary neuroscience. Much of the originality of the book lies in the fact that this has not been done before… [the] position is – as far as I know – unique in contemporary debates in the philosophy of mind.\textsuperscript{99} But, and this is the point that will be argued for in this work, Davidson and von Wright can be seen as carrying on the same agenda as Bennett and Hacker, only not with the same detailed view about neuroscience. In fact, many of the arguments of Bennett and Hacker bear a resemblance to the ones von Wright puts forward in his \textit{In the Shadow of Descartes}.\textsuperscript{100} This is obviously not a surprise since there is a common figure – Wittgenstein – behind these philosophers. Insofar as the position of Bennett and Hacker is not especially original but anticipated by the works of Davidson and von Wright, the views of these latter philosophers are brought into the middle of a recent discussion. This undermines Hacker’s view that a Davidsonian philosophy of mind is anti-Wittgensteinian. On the contrary, it can be used to defend the kind of position which Hacker himself argues for. Very few philosophers have paid attention to this side of Davidson.\textsuperscript{101}

My purpose is not to provide a detailed description of Wittgenstein’s views but to show how a Wittgensteinian attitude, the exact nature of which is itself a matter of serious on-going dispute, helps to explain how Davidson and von Wright can answer to the challenge posed to the philosophy of mind and why an answer to this challenge is important.\textsuperscript{102} Wittgenstein’s conception of philosophy clarifies one way in which the eliminative challenge for the philosophy of mind can be treated, and it gives one refreshing and interesting way to think about the philosophical problems of the mind. Without attempting to claim that this is the right or only method, I suggest that it is a method which contemporary philosophy of mind should not neglect and one which deserves more attention. Von Wright thought that it would not be surprising if Wittgenstein’s contributions to the philosophy of psychology would actually be the most lasting contribution of his philosophy.\textsuperscript{103} Wittgenstein’s views are largely neglected in the modern philosophy of mind because they challenge the very way in which much of contemporary philosophy is being done. It is interesting that although Wittgenstein is widely recognized as one of the greatest philosophers of the 20th century, his philosophical insights have been successfully silenced. There is perhaps a very practical

\begin{flushleft}\textsuperscript{99} Searle, 2007, 101. \\
\textsuperscript{100} The book is surprisingly unknown. I am aware of only one review, by Kivinen, 1998. \\
\textsuperscript{101} Philosophers who have seen this side of Davidson include Stoutland in his later writings, perhaps Rorty and Jeff Malpas. \\
\textsuperscript{102} Wittgenstein is such a complex thinker that I do not expect to describe his views here in a way which could not be seriously criticized by Wittgenstein scholars. Whether or not my description of Wittgenstein is correct is not crucial. What is important is the conception of philosophy which can be found from his writings and the insights that can be further developed. \\
\textsuperscript{103} Von Wright, 1995.\end{flushleft}
reason for this. As Hacker notes: “If Wittgenstein’s warnings against emulating the methods and goals of science in philosophy are warranted, then much contemporary work in philosophy of language, philosophy of mind and of action is no more than houses of cards.” 104 Indeed, Anthony Kenny noted already in 1984 that:

[... ] some of the philosophical gains we owe to Wittgenstein seem in danger of being lost. This is not because his work has been superseded... but... some succeeding philosophical genius. Rather, his contribution has been neglected because more and more philosophers, especially in the United States, have attempted to model their studies on the pattern of a rigorously scientific discipline... holding up... an abstract system for artificial intelligence as the goal of philosophy of mind. 105

Interestingly, a similar remark has been made about Davidson. According to Rorty: “Many who have no use for Wittgenstein have none for Davidson, and for the same reason: to adopt the views of either would be to dissolve problems which they have spent the best years of their lives trying to solve.” 106 In the following three sections I clarify Davidson’s and von Wright’s views about the nature of philosophy and discuss how their views relate to a Wittgensteinian view of philosophy. The purpose of this clarification is to show how Davidson and von Wright could answer the challenge.

1.2.1 A Wittgensteinian view of philosophy

How Wittgenstein saw the nature of philosophy is an on-going and severely debated issue among Wittgenstein-scholars. 107 Wittgenstein’s remarks on this question are scattered around his works and it is not easy to reach a final conclusion about their nature. 108 My purpose is to briefly describe a broadly Wittgensteinian view without taking a strong stance on the question whether this was Wittgenstein’s actual view. I claim that a Wittgensteinian attitude can be found in the views of Davidson and von Wright, although these two philosophers are not “Wittgensteinians” in the same sense as certain contemporary philosophers clearly are. 109 I am not defending the view beyond briefly suggesting why it could seem reasonable.

106 Rorty, 2005. The reference is to an electronic review for which page numbers cannot be given. Davidson (1998a, 91) commented that: “Rorty sees some of my views as serving his Wittgensteinian agenda, which is flattering if deserved.”
108 Glock (2007) distinguishes between rationalist and irrationalist interpretations and makes an additional distinction between seven irrationalist interpretations, which include postmodern interpretations and therapeutic interpretations. He also makes a distinction between “extrinsic and intrinsic interests” in Wittgenstein research. The former include, for example sociological, political and psychopathological angles on Wittgenstein.
109 I am thinking such philosophers as Stanley Cavell, Cora Diamond and Marie Mcginn.
In the *Tractatus Logico-Philosophicus* Wittgenstein famously claimed that philosophy is not one of the natural sciences and that philosophy is something which stands above or below, but not beside the natural sciences. This is metaphorical talk, but the distinction between a Wittgensteinian and a Quinean naturalistic conception of philosophy is clear. Hacker has noted that: “If Quine is right, then philosophy is an extension of science…. If Wittgenstein right, then philosophy is *sui generis*.”

Roger Gibson, an important Quine-scholar, perceptively comments that certain philosophers “[…] not without reason view the naturalistic philosopher as having sold out, that is, as having given up being a Philosopher. We are left, then, with two conceptions of philosophy.” This highlights the on-going debate which concerns the self-image of philosophy. If philosophical activity is something else than scientific activity, if the methods and results of philosophy differ from those of science, what then is the real purpose of philosophy? Does philosophy aim at truth, as science does? The answer, according to naturalists, is clear. Philosophy does *not* differ from science except perhaps in the broadness of its questions. There is no “philosophical method” which could challenge the scientific one. Before accepting the truth of this naturalistic view, let us consider Wittgenstein’s remarks about philosophy:

The object of philosophy is the logical clarification of thoughts.

Philosophy is a battle against the bewitchment of our intelligence by means of our language.

Philosophy unties knots in our thinking.

Philosophical investigations: conceptual investigations.

Philosophical investigations try to untie the perplexities that we have created for ourselves, or which have occurred when we have used language too freely. This suggests that philosophical

---

111 Gibson, 1997.
112 Wittgenstein, 1933, 4.112. Compare to: “[…] logical investigation explores the nature of all things. It seeks to see to the bottom of things and is not meant to concern itself whether what actually happens is this or that. It takes its rise, not from an interest in the facts of nature, nor from a need to grasp causal connexions: but from an urge to understand the basis, or essence, of anything empirical” (Wittgenstein, 1953, §89).
problems are, above else, problems of language. Philosophy or philosophizing, the philosophical method, gets its special nature from the nature of philosophical problems. The problems arise when “language goes on holiday” and – because of this – philosophical problems are not empirical ones. Since philosophical problems result from linguistic confusions, they cannot be solved by discovering new facts, by empirically finding out how things really are. Wittgenstein notes: “[...] our considerations could not be scientific ones... and we may not advance any kind of theory.” This kind of view is in sharp contrast with current views in the philosophy of mind where the demand for the naturalization of the mind, i.e. for a naturalistic theory, is pressing. The considerations which aim for a naturalistic theory of are precisely scientific ones.

A problem which has, in principle, a scientific solution is not a “knot” in our thinking in the same sense as a philosophical problem is. As Wittgenstein says, a philosophical problem has the form: “I don’t know my way about”. One does not know even how to begin with the answering attempt. The claim that a philosophical question cannot be answered in scientific terms is one which crucially distinguishes scientific and philosophical questions. This criterion is problematic, because how could it be known in advance which questions have what kind of answers? When one encounters a question which makes one realize that one does not know one’s way about, the reason for this should be recognized. According to Wittgenstein, the way has been lost because words have been detached from their ordinary use and are instead used in way which causes misunderstandings. These misunderstandings must be cleared away by clarifying the logic of our thought. This, in turn, is done by looking into the actual workings of language. Philosophical problems which have occurred because people have started to use language in a strange way, for example by drawing analogies between expressions in different regions of language, are nevertheless not merely linguistic confusions. The problems are not mere word-games. Wittgenstein notes:

The problems arising through a misinterpretation of our forms of language have the character of depth. They are deep disquietudes; their roots are as deep in us as the forms of language and their significance is as great as the importance of our language.

Humans are linguistic creatures, and therefore the importance of language to life is profound. The importance of the problems which arise due to language is as remarkable as the

---

117 Fodor (1994) for example writes that the “naturalization project” is an attempt to show how something – in this case a “mental representation” – can both represent and cause.
118 Wittgenstein, 1953, §123.
119 Wittgenstein, 1953, §111.
importance of language itself, and therefore a solution to these problems is extremely complex. According to Wittgenstein, philosophical problems keep occurring and seem to remain unsolved because ordinary language has remained relatively unchanged. Our understanding receives the same “bumps” over and over again when it runs against the limits of our language. As long as we do not recognize that this is the source of philosophical problems, we continue to suffer from a philosophical illness, because philosophical problems really trouble many of us and we do not know how to get rid of them.

The main difference between a Wittgensteinian conception of philosophy and contemporary naturalism is that whereas the latter approach aims at explanatory theories and suggests that the philosophical problems of the mind are scientific problems, the former suggests that the task of philosophy is not explanation but description. But is philosophy something which should aim at truth in the same sense as science, should philosophy change our knowledge about empirical facts or increase our knowledge with respect to them? Should philosophy be an explanatory project? Wittgenstein clearly thinks that this is not the task of philosophy. The task of philosophy is to give a presentation of the actual use of language and the language-games in which expressions are actually used must be respected. As Wittgenstein notes: “Philosophy may in no way interfere with the actual use of language; it can in the end only describe it… It leaves everything as it is.”

Philosophy leaves everything as it is because once the linguistic confusions have been clarified the original problem ceases to be. If one keeps on craving for a more fundamental answer, one is still under the illusion that the source of the philosophical problem lies outside language. When a philosophical problem dissolves, the problem has not been solved by presenting new information as in the case of science – rather, it has been dissolved by re-arranging what we have always known. What we know and have always known is based on the facts about how words are used in the course of our lives. This, after all, is the bedrock of meaningful communication, of social institutions – in sum, of our lives as we currently understand them. It is also the bedrock of the problems.

What is wrong with the tendency of philosophy becoming or trying to become more scientific? The influence of science on philosophy is not necessarily corruptive, but it should be recognized that if philosophical problems have their peculiar nature, if they are due to linguistic confusions and misuses of language, then science will be a wrong place to look for

---

120 Wittgenstein, 1953, §124. Or as Wittgenstein (1953, §126) also clarifies the task of philosophy: “Philosophy simply puts everything before us, and neither explains nor deduces anything. Since everything lies open to view there is nothing to explain.”
the answers. If philosophy and science are approaches with different purposes and different sets of problems, then one approach should not try to infiltrate the domain of the other. Because the line between philosophical and non-philosophical questions is not sharp and cannot always be drawn, it is understandable that overlaps do occur. But the question is whether philosophy can be entirely replaced by science; that is the naturalistic challenge of science for philosophy, and to this question the answer is no as long as the peculiar nature of philosophical problems is admitted.

How could this insight be applied to problems in the philosophy of mind? Wittgenstein’s claim is that the concepts in terms of which these problems are formulated are concepts whose meaning is determined by the roles that they have everyday life.\textsuperscript{121} He writes: “Psychological concepts are just everyday concepts. They are not concepts newly fashioned by science for its own purpose, as are the concepts of physics and chemistry.”\textsuperscript{122} Psychological concepts, in terms which many philosophical problems about the mind are, or at least have been, formulated, are concepts which are familiar to all. Even eliminative materialists who claim that the concepts of folk-psychology should and will be replaced by the concepts of future neuroscience explain and evaluate their own behavior in terms of everyday concepts. These philosophers believe this and that, have desires and expect that they can rely on the folk-psychological system in their interactions with others. Whether this mode of description will be replaced in the far future is one question; another is its indispensability now. But it is also obvious that we do not command a clear view of the use of the words in terms of which we in any case have to continue speaking. The misuse of these concepts and the confusions about their “correct” use creates deep perplexities, and therefore, or so it could be claimed, the task of philosophy is to clarify the nature of everyday concepts. Philosophy aims to clarify the meaning of these concepts in terms of a descriptive analysis. The use of mental concepts in the course of ordinary life is quite unproblematic. But when extraordinary areas are approached, the use of ordinary concepts becomes problematic. Examples of these kinds of cases could be the application of these concepts to animals or machines or – as modern neuroscience does – to the brain. In these cases, it is not clear any longer how the use of everyday concepts should be understood, because the concepts have been so severely detached from their ordinary usage.

\textsuperscript{121} As is well known, Wittgenstein (1980, §687) claims that “words have meaning only in the stream of life.” The context in which expressions are meaningful is the context of everyday life.\textsuperscript{122} Wittgenstein, 1980, § 62.
The first difficulty of a scientific philosophy of mind is thus its attempt to answer questions which do not and cannot have scientific answers. The problems for which philosophy is a cure are not empirical, in the sense that they could be solved by consulting “mother nature”, by checking how things are, by considering where nature’s joints lie. The problems are outside the domain of natural science. The possibility of scientific philosophy is based on the view that a philosophical puzzlement could be removed by means of a new discovery, that the puzzlement would suddenly go away when we understand in more detail how nature works. Indeed, the reason why there are no real philosophical problems according to scientifically oriented philosophers is precisely this: solutions to “philosophical problems” can be found in the same way as solutions to clearly scientific problems.

The second difficulty is that by looking for answers in the wrong places, a scientific philosophy of the mind which attempts to replace ordinary mental concepts or tries to justify their use – as if something like this were needed – is in danger of becoming wildly speculative by transcending what is actually empirically verifiable. This task, speculation in terms of thought-experiments, is or has traditionally been one task of philosophy. But it should not be the task of scientific philosophy, according to those who think that literal truths can be achieved only through scientific enterprise. The insistence that philosophy should be closely continuous with science leads to a speculative philosophy which has no real grounding. Wittgenstein famously wrote:

> Philosophers constantly see the method of science before their eyes, and are irresistibly tempted to ask and answer questions in the way science does. This tendency is the real source of metaphysics, and leads philosophers into complete darkness.  

These words were written a long time ago but they are very relevant today as the rise of scientific philosophy testifies. The metaphysical attitude which may lead to the assimilation of philosophy and science should be resisted, because, as Stoutland for example has noted, the metaphysical constructions of a scientifically oriented philosopher give the illusion of resolving philosophical problems when in fact the real problems have been replaced by puzzles generated by this metaphysical activity itself. Stoutland, following Wittgenstein’s lead, suggests that a philosopher who gets lost in such metaphysical speculations “[…] is no longer looking at the way concepts work in our life and thought; she is considering how they must work in the context of a metaphysical construction of her own creation.”

---

123 Wittgenstein, 1958, 18.
124 Stoutland, 2006a.
125 Ibid.
important observation. Speculation guides the naturalistic philosophy of mind; naturalization is required for the vindication of mental phenomena because otherwise they would not fit into the framework required by the demands of physicalism. Mental phenomena must be naturalizable because they must be part of the metaphysical framework of physicalism.

The failure to see that philosophical problems arise because language goes on holiday leads to the view that a philosophical problem is something which requires a scientific solution or explanation, and thus to a demand that there must be a solution expressible in scientific terms. Von Wright traces the return of speculative metaphysics, the rehabilitation of the speculative, as he calls it, as being an unfortunate consequence of the rapid development of the empirical sciences. Von Wright takes Chomsky’s “Cartesian linguistics” as a paradigm example of a speculative metaphysics. Also Davidson, in his criticism of Chomsky, Fodor and Pinker on their postulation of the ‘language of thought’ sees that this kind of rehabilitation has occurred. New empirical findings require explanations and a theoretical framework that accommodates these findings. But the construction of a metaphysical framework in order to understand or explain how empirical facts go, or as Stoutland notes’ must go together, results in a situation in which concepts get hopelessly detached from their original and actual use in human life. The metaphysical constructions are nothing but, to use Wittgenstein’s famous expression, houses of cards which will collapse when the concepts that helped to build the castles have been completely isolated from their actual use. What is the point of these construals – a la eliminative materialism – if the concepts from which we started have been transformed and lost beyond recognition? Do the scientific revelations reveal the nature of the original concepts? Whenever a relapse into speculative metaphysics occurs, we should ask: “…is the word ever actually used in this way in the language-game which is its original home?” A battle against speculative and scientistic metaphysics should be fought by bringing words back from their metaphysical to their everyday use.

This brief description of a Wittgensteinian view of the nature of philosophy is meant to suggest a way in which philosophy can answer the challenge posed by science. Philosophy can do this by insisting that philosophical problems differ from empirical problems and by being an activity which differs from the scientific enterprise. An important task of

---

126 Von Wright, 2000b.
127 See Davidson, 1997.
129 Whether “a battle against metaphysics” should be carried out is a question in its own right, which I shall not further consider in this work. Wittgenstein had his own reasons to be suspicious about speculative metaphysics. The question that interests me here is not whether “speculative metaphysics” is something that should be resisted. I merely note that contemporary scientific philosophy is, contrary to its own principles, often wildly speculative.
philosophers is to clarify how philosophical problems differ from scientific ones. This helps us to understand why these problems do not have similar solutions as scientific ones, and eventually dispels the desire to see philosophy and science on a par. Von Wright, who certainly had a good grasp of Wittgenstein both as a person and as a philosopher insofar as these two can be separated, described the Wittgensteinian conception of philosophy as being a *therapeutic conception of philosophy* because its purpose is to “exorcize” or “demolish”, and *not to solve*, the disquietudes of the mind which arise from a failure to understand the logical grammar of language.\(^{130}\) This is one way to see Wittgenstein’s view about the form of philosophical problems and also his view about the nature of philosophical activity, which is very different from the naturalistic view about the nature of philosophy. In the *Tractatus* Wittgenstein states that: “Philosophy is not a theory but an activity”,\(^{131}\) and later this view of philosophy as *a kind of therapy* is developed further. We hear: “The philosopher’s treatment of a question is like the treatment of an illness”,\(^{132}\) but because philosophizing is an activity, perhaps quite a personal one, there is no straightforward answer as to how the philosophical illness is cured; it is instead the case that: “There is not *a* philosophical method, though there are indeed methods, like different therapies.”\(^{133}\) Philosophy is a form of conceptual therapy which heals by appealing to ordinary language. This therapeutic activity is not theoretical or explicative, it is descriptive. It is clear that naturalistic philosophy does not want to have anything to do with the view that philosophy should be understood as “therapy” of any kind.

I have clarified the Wittgensteinian conception of philosophy because from this perspective the challenge of science can be resisted, and because this conception clarifies also the Davidsonian / von Wrightian understanding of philosophy. I suggest that, in important

---

\(^{130}\) Von Wright, 2000. As von Wright (2000b, 281) says, the therapy called philosophy dissolves “linguistic fixations” or “mental cramps”. Our “mind comes to rest” when we cease craving an answer which cannot be had and *realize why this is so and has to be so.*

\(^{131}\) Wittgenstein, 1933, §4.112.

\(^{132}\) Wittgenstein, 1953, §255.

\(^{133}\) Wittgenstein, 1953, §133. Consider also: “In philosophizing we may not *terminate* a disease of thought. It must run its natural course, and *slow cure* is important.” (Wittgenstein, 1967, §382) How this and other cryptic remarks, like references to “illnesses” or “therapies” should be understood is not entirely clear. Although Wittgenstein uses these kinds of expressions when discussing the nature of philosophy he was annoyed, according to von Wright (2000b), with the label “therapeutic positivism” which was often applied to his philosophizing. Perhaps one way to understand the therapeutic conception of philosophy is to think of philosophy as being a very subjective activity in which one is only concerned to clarify one’s thoughts to oneself i.e. to release oneself from the problems which have been tormenting one. If also others benefit from this: good. The benefit will be a new way of seeing something familiar. In Wittgenstein’s case, it is often said that his philosophy cannot really be separated from his life in general and it is clear that “deeper” problems about life certainly occur in his philosophical writings, (e.g. in Wittgenstein, 1961) in a sense which is alien to most contemporary (naturalistic) thinkers. Wittgenstein (1967, §456) has a point when he notes that: “Some philosophers (or whatever you like to call them) suffer from what may be called loss of problems. Then everything seems quite simple to them, no deep problems seems to exist anymore, the world becomes broad and flat and loses all depth, and what they write becomes immeasurably shallow and trivial.”
respects, both Davidson and von Wright understand the nature of philosophy in a similar way to Wittgenstein. In von Wright’s case this is often recognized but not sufficiently emphasized. The connection between Davidson and Wittgenstein has so far remained without discussion.

1.2.2 Davidson’s conception of philosophy

The remarks about Davidson’s conception of philosophy here may be controversial. Whether or not they capture his actual view is a secondary question: my main interest is to show that there is a way in which a Davidsonian could answer the challenge posed by science. I want to show that there are important insights to be gleaned from Davidson’s writings – insights which he did not develop further.

In his first published article entitled “Why Study Philosophy?” Davidson, when considering how the study of philosophy could be justified, writes that most arguments against philosophy: “[…] err because they do not take account… the special character of philosophical activity, and therefore try to give a wrong sort of justification for it.”134 In Davidson’s work, in contrast to that of many philosophers who are not really interested about philosophy, the question about the nature of philosophical activity is thus raised right in the beginning. The worst mistake is to think that in order to be justified, philosophy must be one among the sciences or perhaps even above science, as a sort of super-science. This view, that philosophy is not among the sciences, is related in spirit to the Wittgensteinian conception of philosophy. According to Davidson “philosophy is not among the sciences.” Yet he counts himself as a naturalist and a proponent of the standard Quinean naturalistic view usually declares that: “Naturalistic philosophy is continuous with natural science.”135 This view implies that philosophy is among the sciences, it is not above or below science, as Wittgenstein suggests.136 Two important philosophical figures – Quine and Wittgenstein – each holding different views about the relation between philosophy and science can be seen as influences for Davidson. So what to make of this?

Hacker, for example, makes a firm distinction between a Quinean and a Wittgensteinian view. Von Wright makes a similar distinction when he draws the line between a Russellian and a Wittgensteinian conception of philosophy. How coherent is an amalgamated Quinean–Wittgensteinian view about the nature of philosophy, which Davidson

134 Davidson, 1952, 22.
135 Quine, 1995, 256-257.
136 Strictly speaking “continuous with” does not by itself suggest that philosophy is “among” the sciences. But, as Hacker notes, the Quinean view is that philosophy is an extension of science. This is denied by Davidson.
seems to be accepting? Davidson is a naturalist who accepts the results of common sense and science. But this is not a reductive version of naturalism, because it does not involve the idea—
to use the example from the philosophy of mind—that a naturalization of mental phenomena requires showing, or attempting to show, how mental states might be reduced to something that can be subsumed under the natural sciences. In Davidson’s naturalism, science isn’t privileged in the way it is in the views of reductive and eliminative materialists. But, why isn’t a reductive naturalization required? Why isn’t our mental vocabulary in need of “vindication” or justification? The complete answer to this question will emerge later in this discussion, when the arguments in favor of the irreducibility of the mental will be considered. My interpretation is that the general motivation for resisting reductive naturalizations is the belief that a reductive approach to philosophy would jeopardize the distinctive character of philosophy by bringing philosophical problems closer to scientific ones or even assimilating the two kinds of problems. This would go against Davidson’s conviction that the natures of these problems are essentially different. On the other hand, a reduction of, for example, mental phenomena would try to answer certain kinds of problems in the wrong kind of way. The questions for which one seeks answers “at the mental level” receive the wrong kind of answers if the answers are formulated in physical terms. Whereas Quine and his ideological followers insist that mentality needs “vindication” or that the use of folk-psychological concepts must be “justified” by science, Davidson by contrast takes the Wittgensteinian attitude, according to which it is absurd to require vindication of concepts which are currently indispensable.

Davidson is thus a naturalist who rejects the idea that our mental vocabulary could or should be replaced by the vocabulary of science. As a part of his naturalism, Davidson accepts Quine’s abandonment of the analytic / synthetic distinction. The consequent blurring of the boundary between philosophy and science is treated with suspicion by those philosophers, Wittgenstein among them, who think that a priori conceptual analysis is one

---

137 For Davidson’s understanding of naturalism see Davidson, 1995 and 2000. In the latter, Davidson challenges the view that naturalization requires showing, or trying to show, how to reduce mentality to something which can be understood in non-mental terms.

138 I suggest that Davidson’s position comes close to “classical naturalism” and insists that the study of reality should not be carried on in terms of methodological monism. Davidson’s naturalism, which is sensitive to biological facts about humans and tries to non-reductively explain the existence of thought from this position, comes out in full force in Davidson’s later philosophy when the idea of triangulation is introduced. This is a model the purpose of which is to explain the emergence of thought as a result of our natural capacities and social interactions. This suggests that there is nothing supernatural involved in the process in which thought emerges, and it should be enough to defuse the cravings to “vindicate” mentality through “naturalizations”.

52
defining feature of philosophy or that such analysis is a philosopher’s task.\textsuperscript{139} Although Davidson expresses a commitment to naturalism, he believes that philosophy \textit{can} provide solutions to empirical problems through conceptual analysis. In this sense Davidson’s view differs from Wittgenstein’s, who thinks that when a philosophical problem is analyzed it is eventually realized that the original problem vanishes. The tormenting problem disappears through the realization that the problem was a pseudo-problem. The realization is reached \textit{through description}. This is also von Wright’s view.

Contrary to von Wright and Wittgenstein, Davidson seems to think that philosophy can make an important contribution to problems which are clearly empirical. In his philosophy, Davidson seeks explanations in a different way than Wittgenstein does. But how the results of Davidsonian conceptual analysis should be understood is, in my opinion, unclear. Sometimes Davidson suggests that \textit{the task of philosophy is merely to describe features of certain concepts}\textsuperscript{140} and that he tries “to express what is attractive about a position by showing how it fits together with other things which we find plausible”.\textsuperscript{141} This is a different way of describing how to fit things together than the demand that “higher level phenomena” need naturalization if their “proper place” in the “grand scheme of things” is to be described. It is also a specific view about the nature of philosophical activity, and it does not suggest that philosophy should aim for explanations in the same sense as science – and does not hold that the main task of philosophy is to come up with substantial explanations of phenomena. \textit{The task of a philosopher is to describe features of concepts}; it is concepts that are being analyzed, not “things in themselves”. This strategy can be seen resulting from the Quinean view that the line between questions of fact and questions of meaning cannot be clearly drawn. This blurs the distinction between metaphysics and epistemology. From a Quinean viewpoint, the results of conceptual analysis are treated with suspicion since a naturalist is wary of the concept of the \textit{a priori}. On the other hand, Quine makes a metaphysical distinction in his insistence that some discourses are more “factual” than others. Rorty claims that in Quine’s case the distinction between metaphysics and epistemology is

\begin{flushright}
\textsuperscript{139} Wittgenstein is such a complex thinker it is impossible to conclude what a defining feature of philosophy, if such there is, would be according to him. However, Wittgenstein certainly thinks that the method of philosophy differs from that of the sciences (the term conceptual analysis as a description of the method of philosophy seems not to be out of place here), and whereas philosophy may not discover new facts \textit{philosophizing nevertheless has consequences}.
\end{flushright}

\begin{flushright}
\textsuperscript{140} See, for example, Davidson, 1982. Davidson (1995, 205) notes that this kind of description of concepts could also be called an \textit{analysis of them}. This view is in contrast to the naturalistic Quinean view, according to which there are no conceptual truths and therefore there are no conceptual analyses either. This being said, it should be acknowledged that Quine’s own philosophy was speculative and done from the armchair.
\end{flushright}

\begin{flushright}
\textsuperscript{141} In an interview by Jones, 2003.
\end{flushright}
brought back precisely through science. It seems to me that if Davidson really accepted Quinean naturalism, it would be unclear what his conceptual analyses could actually achieve.

According to Davidson, the philosophical description of concepts is not reductive since it does not attempt to explain concepts in terms of other concepts that are thought to be more basic or clear. The way I read this is that the most basic concepts are without foundation, which would transcend the way that these concepts are used. The foundation of basic concepts lies in the ordinary language. Although philosophy aims at the clarification of concepts, it does not aim at definitions. The purpose is not to lay down rules that would show how concepts should be used. A philosopher is not an expert in forming definitions, but he is perhaps skilled enough to clarify the nature of already existing concepts in cases where confusions occur. Following a Wittgensteinian view, it could be claimed that this is a normative claim about philosophy and the task of philosophers. Philosophers should focus on the clarification of concepts in cases of occurring confusions; they should be skilled enough to do this. Of course, there cannot be any “proof” that this is how things should be. Instead, my suggestion can be read as a description how to think about philosophy and how certain philosophers, for example Davidson and Wittgenstein, see their own conceptions of it. Clarification is the main task of a philosopher. In an important passage in which Davidson describes his method, we read:

In philosophy we are used to definitions, analyses, reductions. Typically these are intended to carry us from concepts [that are] better understood, or clear, or more basic epistemologically or ontologically, to others we want to understand. The method I have suggested fits none of these categories. I have proposed a looser relation between [the] concepts to be illuminated and the relatively more basic.

Davidson claims that clarification, not in terms of reduction or explicit analysis but in terms of understanding, teaches us more than the efforts to produce correct and revealing definitions

---

143 Also this goes contrary to the naturalistic view that mental concepts should be analyzed in terms of physical concepts because the latter are more natural or clear. The demand for naturalizations is based on this view.
144 Davidson, 1973a, 136. See also Davidson, 1977, 219. With regard to irreducible basic concepts, like truth, knowledge, belief or action, it should not be expected that they could be reduced to more basic concepts, but something revealing about them can be said by relating them to other concepts. What makes a concept irreducible is less than clear. Davidson (1998, 85) only notes that a concept is irreducible if it cannot be defined in terms that are as general and at least as clear as the concept to be reduced. In my opinion, the questions of what concept is clearer or more general than some other concept and how concepts can be compared in terms of their generality is problematic and something for which Davidson does not propose a solution. A basic or irreducible concept is a concept that cannot be defined in terms of other concepts. Why, according to Davidson, such definition is unavailable is unclear. Philosophy should not aim at definitions, but a basic concept is nonetheless defined in terms of irreducibility.
or reductions of basic concepts in terms of clearer or more fundamental concepts.\textsuperscript{145} Why? Because reductive analyses of the most primitive concepts have always failed, whereas philosophical clarifications of the kind to which Davidson refers have already revealed something. This of course is merely Davidson’s view, and to claim that so far there has been a systematic failure in reductive analyses is perhaps not a very compelling reason to believe that such analyses are impossible. Yet, a suggestion that a philosopher tries to clarify in terms of understanding can be seen again as a suggestion of the nature of the philosopher’s task. As we will see, this view is very close to von Wright’s view of the task of the philosopher. Clearly the question of which kind of strategy “teaches us more”, or which kind of attitude towards philosophical questions should be chosen, is once again something for which there cannot be an answer which would remain beyond dispute. In what sense philosophical clarifications have actually increased our understanding or explained anything is also a question the answer to which must be left for everyone’s own intuitions, and for which I shall not advance an answer in this work. From a Wittgensteinian perspective, it could be said that a philosophical clarification has occurred insofar as one has been cured from a particular philosophical disease or from the disease of philosophy in general. Whether the latter would lead to the end of philosophy is a question worth considering.

What is important in Davidson’s way of doing philosophy is the attempt to see things in a new light and to clarify conceptual muddles. It is not surprising if this way of putting things sounds metaphorical. After all, it cannot be clearly described or explained what happens when a person \textit{understands} something which he did not understand before – i.e. it is not possible to describe in mental, not to mention physical, terms what “seeing things in a new light” means beyond what can be intuitively grasped about the meaning of this metaphorical expression. Wittgenstein once wrote that: “Thoughts that are at peace. That’s what someone who philosophizes yearns for.”\textsuperscript{146} This is a beautiful description of how philosophy can benefit an individual; perhaps seeing things in a new light could be understood simply as a process which brings peace to thoughts so that philosophical problems do not trouble one anymore. Notwithstanding how we understand the nature of philosophical clarification, the philosopher’s task in Davidson’s view, as in Wittgenstein’s case also, is \textit{to re-arrange what we already know instead of coming up with new empirical knowledge}. The puzzles often result from a conflict between things that one already knows, and therefore a

\textsuperscript{145} A claim according to which we try to \textit{understand} basic concepts in terms of others is vague, as Davidson (1977, 219) himself admits.

\textsuperscript{146} Wittgenstein, 1980a, 43.
solution should not look beyond what is already familiar. When discussing the work of Einstein, Davidson notes that the philosophical activity of Einstein was the “[...] clarification of basic concepts”\textsuperscript{147} and that in this case philosophy helped by “[...] providing a new and more productive way of thinking about empirical concepts”.\textsuperscript{148} Whether this is the “right” way to do philosophy is not an essential question since, as noted already, the idea that philosophy would have only one purpose or that there would be only one correct philosophical method is utterly implausible.\textsuperscript{149} I merely want to emphasize that Davidson has a certain view about the nature of philosophy and his philosophical arguments should, or better, could be understood against this background.

As the title of Davidson’s first article – “Why Study Philosophy?” – shows, he seems to suggest that philosophers should contemplate the question of what they are doing when they practice philosophy. It can be said, at the very least, that this question interested Davidson. One way to understand the Wittgensteinian conception of philosophy is to see philosophy as a form of therapy which, instead of solving its problems, exorcises them. It could be claimed that exorcising is solving; perhaps the appropriate term for a philosophical solution would therefore be therapeutic solution. Interestingly, what Davidson says about the nature of philosophy resembles this view. He writes:

\[
\text{[...] the disease for which philosophy is a specific cure is not confusion… but the anxiety of confusion. If you do not mind being vague and muddled, no one can… promise that you will be better off less so. But if you do mind, then you have a philosophic disease, and the study of philosophy can help you make feel better. It can loosen your mind muscles, smooth out the dogmatic cramps, prescribe exercises which will be a step toward greater intellectual ease.}\textsuperscript{150}
\]

These lines bear a curious resemblance to von Wright’s and Wittgenstein’s views. Both von Wright and Davidson talk about “cramps”, which philosophy is meant to relax. Whereas Wittgenstein is yearning for “thoughts that are at peace”, Davidson is looking for “intellectual ease”. Perhaps we could, programmatically, suggest that philosophy is best suited for those who, unfortunately, are troubled by questions which do not trouble most people; Anthony Kenny has claimed that philosophy is useful only for sick people and Wittgenstein thought that a philosopher must cure himself before helping others.

\textsuperscript{147} Davidson, 1952, 23. Davidson refers to Einstein because, although being a physicist, he was by his own account engaged in a philosophical enterprise.

\textsuperscript{148} Ibid. Emphasis mine.

\textsuperscript{149} There are correct methods in science, but there one already has a view of what the possible answer could be like, one roughly knows what to look for. In philosophy one does know one’s way about.

\textsuperscript{150} Davidson, 1952, 24.
Davidson claimed that *philosophy is the only hope* for those who are troubled, for example, by the apparent conflict between free will and a deterministic universe. Since philosophy is not among the sciences and philosophy is the *only* hope for those troubled by such conflicts, we can assume that, according to Davidson, this conflict and others similar to it cannot be solved empirically. This understanding of the nature of philosophy is very different, and I would say *deeper*, than our current naturalistic understanding. An analysis of Davidson’s philosophy which ignores this aspect does injustice to his views. It needs to be acknowledged that the goal of a naturalistic philosopher is to find out how certain things are, whereas the goal of a Wittgensteinian–Davidsonian–von Wrightian is to dispel the confusions which exist between ordinary concepts.151 We have seen that philosophers of mind, for example, Kim and Block, have differing views about the question of how empirical the philosophy of mind should be. The differences in views about the *nature of philosophy* are much deeper. A Wittgensteinian–Davidsonian–von Wrightian conception, which sees philosophy loosening the mind’s muscles and offering intellectual ease by providing new and productive ways to think is something fundamentally different from Fodor’s, who wants to integrate philosophy into the scientific enterprise, or Searle’s, who thinks that the purpose of philosophy is to make way for science.

According to Davidson, philosophy can make you feel better if you have a philosophic disease: it can examine, clarify, reconcile, criticize, regroup or unearth our basic convictions and assumptions, but it can also do something more. Philosophy can lead to knowledge which was not in sight at the start of philosophical contemplation and which was not necessarily implicit in what was known before the contemplation. There can be progress in philosophy, but the new knowledge that may be achieved is not empirical knowledge in the sense that it would be based on new facts which are found. It is, rather, a result of *understanding in a new way* what we already knew. The new knowledge is thus a result of conceptual – philosophical – investigations. Where Davidson’s view differs from that of Wittgenstein is that the former is, at least occasionally, trying to give explanations whereas the latter insists that only descriptions must be given. A Davidsonian could perhaps think that there can be progress in philosophy, whereas both Wittgenstein and von Wright are skeptical of this view.

151 Consider Davidson’s (1993, 179, my emphasis) comment: “No one can object to the attack on confusion, conflict, obscurity and self-deceit *in our everyday beliefs*: these defects in our views of ourselves and the world exist in profusion, and if some philosophers can with skill or luck do something about reducing them, those philosophers deserve our respect and support.” The purpose of naturalistic philosophy is not this. With respect to “everyday beliefs”, eliminative materialism creates more confusion than it dispels.
So far I have claimed that Davidson’s conception of philosophy is close to Wittgenstein’s view, and alien to a naturalistic understanding according to which philosophy and science stand on the same level. Yet, it is also clear that Davidson is, some would say very strongly, influenced by Quine’s naturalism. Certain commentators see Davidson continuing the strict physicalistic ideology starting from Quine. But, as Davidson’s non-physicalistic philosophy of mind testifies, this interpretation should be challenged. My purpose is to argue that Davidson’s philosophy of mind is meant to be a novel contribution for physicalism, but that the non-reductive side of his philosophy of mind is more interesting and prevalent than his contribution for physicalism. Perhaps it would be best to say that Davidson’s approach differs from the approaches of Quine and Wittgenstein in an interesting way. According to the Wittgensteinian view, philosophy should not try to provide conceptual truths, because the method of philosophy is destructive (although in a therapeutic way). The method and aim of philosophy is “destructive” because it dispels the problems in question. Philosophy leaves everything as it is. Wittgenstein thought that, because of the nature of language, there are no fixed meanings, and what one says about things is intertwined with issues of an arbitrary nature – above all else, with the actual use of language in the practices of the language-users. Quine, on the other hand, argued that the distinction between claims which are true in virtue of empirical content and claims which are true because of the meaning of the words of those claims cannot be maintained – and this, among other reasons, was why he thought that the boundary between philosophy and science vanishes. Philosophy cannot provide conceptual truths, and is thus on a par with science which does not provide conceptual truths either.

Davidson praises Quine’s conclusions about the hopelessness of the analytic / synthetic distinction but has a Wittgensteinian view about the nature of language, and his way of doing philosophy and achieving philosophical conclusions gives the impression that important truths can be achieved through conceptual analysis. Many of Davidson’s most important philosophical conclusions are not based on empirical evidence. This is especially true with respect to his philosophy of mind, and the a priori style of Davidson’s philosophy has been criticized by many. Fodor notes that Davidson holds a small number of general principles, the application of which allegedly solves a vast number of philosophical puzzles, but concludes that these stunning arguments “don’t quite work.” Hans-Johann Glock, a recent commentator of Davidson’s work concludes that his “a priori arguments are more

---

152 See for example, Mulhall, 1987 and Glock, 2003.
153 Fodor, 2002, 12.
suggestive than compelling”\(^{154}\) and Nannini, another recent commentator, claims that Davidson:

\[
\ldots \text{like many ‘classical’ philosophers (e.g. L. Wittgenstein, G. Ryle and also the logical empiricists), overtrusts the ability of philosophy to solve (or dissolve) a priori, by means of simple conceptual analysis, theoretic problems that arise from empirical sciences or common sense.}^{155}
\]

I think the commentators are right in their claim that Davidson tries to “prove”, in some sense of the word, various things by means of conceptual analysis. Perhaps it is also correct to say that his arguments are more suggestive than compelling, but in my view this is a meritorious aspect of those arguments. After all, one purpose of philosophy should, or could, be to provide new insights, new ways of thinking, and not to compel people to adopt certain views.\(^{156}\) This, obviously, is again only one suggestion or intuition concerning the task of a philosopher. I am not defending it here, but merely proposing its possible importance.

I think, however, that the critics err in their claim that Davidson fails to prove what he intends. This is because they do not take seriously enough Davidson’s view that he is merely describing the features of concepts. This means that Davidson is not even trying to say something about the world i.e. he is not, pace Nannini’s claim, trying to solve problems which arise from the empirical sciences. Wittgenstein was not interested in solving or dissolving empirical problems; the same observation applies to Davidson. Indeed, Davidson makes a firm distinction between the enterprises of the scientist and the philosopher. The distinction follows from the conviction that was discussed in relation to Wittgenstein, namely that philosophical and scientific questions differ in important respects and require different kinds of treatment. Davidson notes, for example:

\[
\text{How many… speakers of a language must there be if anyone can be said to speak or understand a language? Since this is a matter governed by the crooked course of evolution, I have no idea what the answer is; perhaps it takes quite a crowd. But as philosophers we can ask the question in a more theoretical vein.}^{157}
\]

When Davidson describes his approach he notes: “I am arguing, as philosophers will on a priori grounds, that psychology and the social sciences are impossible.”\(^{158}\) According to him

\(^{154}\) Glock. 2003a, 351.

\(^{155}\) Nannini, 1999, 113.

\(^{156}\) What, after all, is a compelling philosophical argument? Is it such that the conclusion must be accepted, that the conclusion cannot be avoided? The compulsiveness of the argument depends on whether or not one accepts the premises. As we have seen, it is often the case that philosophers do not agree about the premises. How could an argument be compelling then?

\(^{157}\) Davidson, 1992, 107, my emphasis.

\(^{158}\) Davidson, 1964, 47, second emphasis mine.
philosophers use a priori arguments which, as we have seen, is a strategy certainly not favored among scientists or naturalists. As far as Davidson’s example above is concerned, I think a naturalist would say that empirical research is required to settle the question of whether language is a social phenomenon – and that the scientific verdict, so to speak, ends the discussion. A naturalist would insist that challenging valid empirical results would be useless philosophical speculation or wordplay. Davidson suggests, however, that philosophers may approach the same questions as scientists do, but philosophers approach them in a more theoretical vein because they are actually engaged “only” in a conceptual exercise. But what is characteristic of theoretical approach favored by philosophers? A partial answer is given when Davidson notes that: “I am not concerned with the scientific explanation of the existence of thought; my interest is in what makes it possible.” Likewise, we see how Davidson is trying to show, by using a priori arguments, that something - in this case social science - is impossible. This is an interesting approach, since it suggests that even though scientists may come up with a scientific explanation for the existence of thought, this would not necessarily satisfy a philosopher. What matters to a philosopher is not the kind of explanation that various sciences can and may provide, but an analysis of the concept of thought, because she is interested of the Kantian-like question of what makes thought possible. In Davidson’s view, the question of what the conditions on the possibility of thought are is not empirical. What characterizes the philosopher’s efforts is a desire to conceptually explore the limits of possibilities.

159 Fodor (2007, 114) for example writes : “I hate a priori arguments that such and such discourse can’t be naturalized...” Paul Churchland (2007a) dislikes the a priori knowledge typical of modern analytic philosophy because it is “factually empty” or “conventional”. It seems to me that philosophy of mind, which uses a priori arguments, will not receive fair consideration from those who “hate” a priori arguments.

160 Davidson, 1995, 208, my emphasis.

161 The question “Why there is such thing as thought?” is used here as an example of a philosophical problem due to the fact that it was a problem which was central in Davidson’s later philosophy of mind. For the ease of exposition I will use this problem as an example.

162 For a reference to Kant on this issue, see Davidson 1995. Consider also: “Of course, we knew it [interpretation] was possible in advance. The philosophical question was, what makes it possible?” (Davidson, 1990, 325). A distinction between an empirical question and a philosophical one is made again when Davidson (1994) notes that he is not interested in what does happen when people interpret each other but what could happen. This means that he wants to give one suggestion about how interpretation is possible. Davidson (1999g, 300) also comments: “...even if I did know [how we actually understand each other], I would still be interested in the question what the formal properties of the system are that make it possible.” This should be compared to Wittgenstein’s (1953, §125) remark about the nature of philosophy: “One might also give the name ‘philosophy’ to what is possible before all new discoveries and inventions.” Consider also: “In a scientific investigation we say all sorts of things, we make many statements whose function in the investigation we don’t understand.... We move through conventional thought patterns, automatically perform transitions from one thought to another according to the forms we have learned. And then finally we must sort through what we have said. We have made quite a few useless, even counterproductive motions and now we must clarify our movements of thought philosophically.” (Wittgenstein, 1980, §155)
I must note that the nature of the “conditions” that Davidson is looking for is nevertheless unclear. Sometimes his purpose seems to be to describe sufficient conditions for various phenomena, while at times he is after necessary conditions. Given Davidson’s skepticism towards modal notions and his rejection of the analytic / synthetic distinction, it is not absolutely clear how these expressions should be understood: for example, it is unclear what the status of Davidsonian necessity is. But it is clear that, when it comes to questions about the nature of thought, Davidson is interested in the philosophical questions, which in his view are not empirical. They could be understood also as empirical questions if one would be interested in finding out why or how, in the course of evolution, creatures with the capacity to think have evolved. But, according to Davidson, this question is “[…] a matter for the speculation or discovery of scientists”, and he makes it very clear that this is not the question which he is interested in. In Davidson’s view, a philosopher is trying to understand the question of what makes thought in principle possible, and his interest is therefore not on the same level as the scientist’s, who is interested in empirical facts. In spirit, this is reminiscent of Wittgenstein, who states that:

It is all one to me whether or not the typical western scientist understands or appreciates my work, since he will not in any case understand the spirit in which I write… I am not aiming at the same target as the scientists and my way of thinking is different from theirs.

This is a telling and an extremely important passage from Wittgenstein, which encapsulates a part of the spirit of his work. Of course I do not want to claim that Davidson would have had subscribed to this view, although it is clear that he too was not aiming at the same target as the scientists. On many occasions Davidson notes that Chomsky, for example, has mistakenly thought that Davidson is providing views which would compete with Chomsky’s claims. But Davidson is not interested in the same phenomena as Chomsky, and his aims and interests are different. Many modern philosophers are interested in the same questions as the scientists; although they are called philosophers, the empirical element in their work is so notable that the question of how their work differs from scientific work is worth asking. As Wittgenstein

---

163 Then the possibility-question would be answered simply by describing how creatures with thought have actually evolved.
164 Davidson, 1995, 207, my emphasis. It is doubtful whether a scientific answer to the question of why thought has evolved can be given. This is not a question which can be answered in evolutionary terms. Given the fact that thought has evolved “naturally” without human intervention, it would be most accurate to say that a question “why thought has evolved” is meaningless. There is a “how” question to be answered by scientists, but there is no reason for thought’s occurrence, and thus the “why” question exceeds the limits of science, or better, the “why” question is not a scientific question unless one believes that evolution has a purpose.
165 Wittgenstein, 1980a, 7, my emphasis.
166 See, for example, Davidson, 1995a.
notes, *his way of thinking* is different from the scientist’s way of thinking. Can the same be said of a contemporary philosopher of mind, say Fodor, who in his own words is “not really very interested in philosophy” and yet counts himself a philosopher?

It is no easy to draw a clear line between an empirical and a philosophical question, and perhaps the line between a philosopher's and scientist's task will remain – and should remain – vague as well. That philosophical and scientific questions form a *continuum* is the point stressed by naturalists. We can imagine that they could ask: isn’t a scientific explanation of the existence of thought *also* an explanation of how this phenomenon is possible? In Davidson’s case this cannot be so because, as I have described, on many occasions he emphasizes that he is not interested in the same questions as the scientists and that the question in which he is interested cannot be solved through scientific explanations. It is also important to notice that, according to Davidson, his approach is “speculative” and does *not* explain the emergence of thought. But it would be strange to criticize Davidson’s position as being insufficiently explanatory, as naturalists would, since his whole point is that a certain kind of explanation cannot be given. The hopelessness of coming up with an explanation about how thought has emerged was something that Davidson took seriously; he noted that he was glad to be a philosopher instead of a developmental psychologist, because the latter may face insuperable *conceptual difficulties* in their explanatory task.

The distinction between a scientific and a philosophical question is emphasized when Davidson notes:

> I do not think I have ever conflated the (empirical) question [of] how we actually go about understanding a speaker with the (philosophical) question [of] what is necessary and sufficient for such understanding. I have focused on the latter question… because I think it brings out the philosophically important aspects of communication while the former tempts us to speculate about arcane empirical matters that neither philosophers nor psychologists know much about.167

It is telling that in Davidson’s view empirical matters are *arcane* while the aspects which philosophy can bring into light are necessary. Certain aspects of phenomena are *philosophically important*. This view differs from the naturalistic attitude, which states that “philosophical” problems require answers in scientific terms. We should note that, according to Davidson, an *empirical* question would lead to *speculation* about arcane matters. This claim is the opposite of the naturalist view. Those questions – scientific ones – which lead to empirical speculations are not candidates for “reasonable philosophic speculation”, as

---

167 Davidson, 1994b, 3. It should be noticed that here Davidson mentions both necessary and sufficient conditions as a target for his philosophical investigations. For a separation between a scientific and a philosophical question see also Davidson, 1982 and, 1995.
Davidson argues. The view that an empirical question gives raise to speculations bears some resemblance to von Wright’s view that a rehabilitation of the speculative has occurred in modern philosophy. Empirical results lead to metaphysical speculations, and it is precisely empirically oriented philosophers who are guilty of these kinds of speculations.

When considering Davidson’s approach, we should ask what the relevance of philosophical speculations or conceptual analyses is. An answer to this question distinguishes Davidson’s approach from those of Quine and Wittgenstein. Davidson agrees with Wittgenstein that a philosopher is interested in different questions than the scientist and the philosophical method differs from the scientific one. But the results of philosophy do not concern only the features of concepts because – and this is distinctive of Davidson’s position – there cannot be a major distinction between the ways one conceptually interprets the world and how it is actually “put together”. Conceptual investigations can lead to substantial results given that language mirrors reality. From a Davidsonian point of view, philosophical conclusions can be seen as empirical or a priori. There is always some empirical element in reasonable philosophical speculation; usually the starting point is based on facts that are already known. In Davidson’s case – as in von Wright’s and in Wittgenstein’s as well – the starting point that must be respected, the empirical element, is the language of the everyday. This language, and what is expressed in it, are more familiar to us than the results of science. The facts of everyday life are something which we all know, but they can become a source of philosophical problems when we let ourselves become confused in the ways Wittgenstein warned about. We become confused when we lose sight of the ways that concepts are actually used; the way they are used is of course an empirical issue.

Davidson’s non-reductive naturalism is sensitive to science and common-sense. Many of the philosophical arguments that he puts forward depend in part “not on purely a priori considerations but rather on a view of the way people are” and “Logic does not assure us of

---

168 See Davidson, 1991, 157. “Reasonable philosophic speculation” is of course a very obscure term, the meaning of which remains without clarification in Davidson’s writings.

169 This is the famous point of Davidson’s classic article “On the Very Idea of a Conceptual Scheme” dating back to 1974. Later Davidson has elaborated the idea that concepts describe truly language-independent reality and that sense cannot be made of the claim that the conceptual structure we employ might not describe objective reality. For this claim, see Davidson, 1984a, 1993a, 1991 and 2000. In these later articles, Davidson’s version of realism is strongly connected to his idea of triangulation, which establishes what the content of observation sentences is.

170 A detailed discussion of this “mirroring” would require a clarification of questions relating to Davidsonian epistemology. This topic cannot be discussed in this work.

171 Davidson, 1999, 194, my emphasis. This, “the way people are”, is not a scientific fact but part of the common knowledge that we all possess. The conclusions of philosophy, say about the emergence of thought, should be understood in the context of our biological nature. Davidson’s philosophical arguments are not based on the way
this [that communication depends on the shared propensities of people to react similarly to their environment], it’s an empirical fact about people.”¹⁷² I think that the role these empirical facts play bears a resemblance to Kant’s synthetic a priori propositions, which are empirical but universal – given human natural history. Certain facts of nature apply to all human beings, but this is not to say that these facts could not be different. Davidson’s philosophical approach has been interpreted as a form of “philosophical anthropology”, which attempts to describe human behavior and human capacities in general and language and linguistic capacities in particular.¹⁷³ Consequently, the cornerstones of his philosophy are philosophy of language, mind and action. Simon Evnine even claims, with some justice, that Davidson’s “technical work in logic and formal semantics, and his metaphysical work on the nature of events and causation… is primarily… apparatus necessary for the completion of his anthropological project.”¹⁷⁴ Glock, another commentator, claims that the guiding question in this project is “What it is to understand other human beings”.¹⁷⁵ I think the project can be broadened to include the question of how the human position could be understood. Incidentally, Wittgenstein also showed an interest in the natural history of humankind. In his own words he was: “contributing… observations which no one has doubted, but which have escaped remark only because they are always before our eyes.”¹⁷⁶ Glock notes perspicuously that Wittgenstein’s investigations into human nature were meant to show that “a change in contingent natural conditions would render plausible or useful concepts and practices other than our current ones, thereby dispelling the appearance that the latter are metaphysically necessary.”¹⁷⁷ Wittgenstein’s purpose was to show that if philosophical problems can be dissolved into uncontroversial facts of human natural history, then a troubled mind should accept this result and achieve peace.

Let us suppose, along Wittgensteinian lines, that it is possible to defend the view that our current concepts are not metaphysically necessary, although they can be a priori in the sense that they apply universally. Let us in other words suppose, as Davidson does, that empirical facts can grant an “a priori” status to certain claims. Here a tempting question is: How strongly can the conclusions of a philosopher conflict with the actual use of language? To what extent must the empirical facts of common-sense be respected? Sometimes, after all,
it is these very facts which lead us astray. Davidson does not offer clear answers to these questions, so I will offer an answer when discussing von Wright’s understanding of philosophy in the next section. Here I want to point out that those, like Fodor, accusing Davidson for the use of *a priori* arguments, ignore the way in which Davidson himself sees the results of philosophy. It is interesting that whereas some, like Fodor or Nannini, criticize Davidson for being *overtly philosophical* because he uses *a priori* arguments, others like Pagin, criticize Davidson for indulging *unjustified empirical speculations* with no direct philosophical significance. These are curious accusations given Davidson’s own conviction that he is interested only in the philosophical aspects of phenomena.

The status of philosophical conclusions is clarified when Davidson, defending one of his philosophical arguments, explains that:

There cannot be said to be proof of this claim [that aspects of our interactions with others and the world are partly constitutive of what we mean and think]. Its plausibility depends on a conviction which can seem either empirical or *a priori*; a conviction that this is a fact about what sort of creatures we are. Empirical if you think it just happens to be true of us that this is how we come to be able to speak and think about the world; *a priori* if you think, as I tend to do, that this is part of what we mean when we talk of thinking and speaking. After all, the notions of speaking and thinking are ours.

The quote explains what the status of *a priori* conclusions based on conceptual investigations is. The philosophical convictions are “*a priori*” *in the sense* that – since concepts are of our own making and we can, in some sense of the word, *decide* how they are correctly used – the applicability of concepts is answerable to *our understanding* of their meaning. Once the application conditions of a concept are in place, there exist “*a priori*” conceptual connections which cannot be violated without turning the concept into another concept, or ultimately into nonsense. Bennett and Hacker, analyzing the conceptual confusions inherent in modern neuroscience, have noted: “It is an empirical fact… that a given vocable or inscription is… used in a certain way in a given linguistic community. It is not an empirical fact that a word, meaning what it does, has the conceptual connections, compatibilities, and incompatibilities that it does.” This is their modern Wittgensteinian view, which agrees with the Davidsonian conception. The “*a priori*” status of Davidson’s claims gets is force from the conviction that ordinary language is a “transcendental ground” beyond which there is no way to go. This is something strongly emphasized also by von Wright; *our* notions must be respected.

178 For Fodor’s critique of philosophy as conceptual analysis see, Fodor, 2000, 2004. For his criticism of Davidson, see, Fodor, 2002.
180 Davidson, 2001d, 294.
181 Bennett and Hacker, 2007, 147.
This being said, it must be acknowledged that concepts can change as a result of non-conceptual changes. Sometimes new scientific results dramatically change the way in which reality is seen. But in many cases the resulting situation is one in which we have to choose how to best describe the situation. Should a change in the applicability of concepts occur, the empirical starting point would be different from the one we would have, had such a change not occurred. As Davidson claims, the way philosophy can help a scientist is by providing a new and more productive way of thinking about empirical concepts. Also this is something that has been recently emphasized by Bennett and Hacker, who claim that the way philosophy can assist science is “not by offering scientists empirical theories in place of their own, but by clarifying the conceptual structure they invoke”.\textsuperscript{182} This is a different view of philosophy than the one shared among naturalists; they offer theories or claim that the task of philosophy is to make way for science. Insofar as philosophy is required to be beneficial, I believe that conceptual clarification is certainly a contribution that philosophy can make. It is however a different kind of contribution from the contribution of science, which describes how things are. Philosophy, on a Davidsonian account, does not aim at knowledge but can guide the way to how the results of science are interpreted and how the consequences of these results are evaluated. A discovery, a scientific claim, that consciousness takes place in a certain part of the brain or that a neural correlate of consciousness has been found does not, in itself, force the conclusion that the location of consciousness has been found. The correctness of the discovery will be decided on the grounds of how well it fits with other things we find plausible.

I would like to suggest, tentatively, that a distinction needs to be made between the purpose of philosophy – between the contributions that philosophy can make – and the nature of philosophical arguments themselves. The contribution that philosophy or philosophers can make, and this is obviously only one opinion, is the one described by Hacker: to clarify conceptual structures. This view is based on the idea that philosophy is not a contribution to human knowledge but a contribution to human understanding.\textsuperscript{183} The justification of this kind of philosophy is in its promise to eradicate the diseases of intellect which trouble us in the form of conceptual confusions. According to Hacker, these kinds of confusions are manifold, extensive and have far-reaching consequences not only in philosophy but also in the various sciences and in our culture at large. These confusions distort our vision of what we really

\textsuperscript{182} Bennett and Hacker, 2007, 128.
\textsuperscript{183} As noted before, “human understanding” is an expression the nature of which has to remain unclear.
know.\textsuperscript{184} Often these confusions have wide consequences, and therefore the kind of philosophy which attempts to clear us of these confusions should be an approach worth continuing.

On a Davidsonian account, philosophy is an activity which attempts to clear of confusions; this in essence is a Wittgensteinian view. But how, in this context, should the status of philosophical arguments be understood? Davidson’s ambitious attempt to prove something with the use of \textit{a priori} arguments is perhaps alien to a Wittgensteinian conception of philosophy, which aims for description. Davidson’s \textit{a priori} arguments, \textit{when they are not meant to clarify conceptual confusions}, can be criticized as being in conflict with empirical knowledge or not taking such knowledge sufficiently into account. Perhaps the criticism that Davidson overtrusts philosophy’s ability to \textit{solve} problems is justified. One should nevertheless acknowledge that since Davidson is not aiming for “definitions” but for a description of the features of concepts, there is no need to argue or defend the view that the results of philosophy would amount to “necessary truths” or other such modally dubious consequences. For naturalists who are suspicious of \textit{a priori} arguments Davidson has a reply: “Since we are thinkers, and it is we who devise the ways of talking and explaining, I am not abashed when I am told that for me the difference between mental and physical is ‘metaphysical’ or ‘a priori’.”\textsuperscript{185} The acceptance of “metaphysical” or a “priori arguments” could be seen just as an expression of the rejection of the analytic / synthetic distinction.\textsuperscript{186} This would undermine the importance and distinctiveness of philosophical conclusions because they would not differ from the conclusions of science. However, given that Davidson thinks that he is providing \textit{necessary} conditions for phenomena, the strength of his philosophical conclusions must, in my view, be somehow \textit{greater} than those of science (which are based on arcane empirical matters). Scientists do not ordinarily make claims about necessity or claim that something is impossible.\textsuperscript{187} If Davidson intends to really describe necessary conditions, this would grant his philosophical conclusions a true \textit{a priori} status.

I nevertheless hesitate to say anything more about Davidson’s conclusions than what has been claimed already. The reason for this is the inconsistencies that can be detected in Davidson’s views when he considers the relevance of his conceptual investigations. Sometimes he claims that conceptual investigations have not much to do with the way things

\textsuperscript{185} Davidson, 1999a, 620.
\textsuperscript{186} See Davidson, 2001d.
\textsuperscript{187} Of course it is often claimed also among scientists that something is not possible. But usually this means something entirely different than, for example, Davidson’s claim that scientific psychology is an impossibility.
actually are. Naturalists aim for substantive explanations, which describe how things are. From this perspective the charge against Davidson is the question: what is the point of these conceptual exercises? My answer is that conceptual exercises ease the diseases of intellect; such conceptual exercises bear a resemblance to Wittgensteinian therapy. In my view, the attempt to clarify conceptual muddles (and not the attempt to say how things are) could be the real importance of Davidson’s philosophy. On the other hand, Davidson suggests that conceptual investigations can reveal important features of, for example, thought and throw indirect light on the actual empirical questions. This is in conflict with other views that he puts forward.

I cannot provide a final solution to the inconsistencies found in Davidson’s conception of philosophy. I think we can nevertheless summarize Davidson’s position by noting that the clarification of concepts is an important task, and this is done by re-arranging the already familiar field of concepts. If a clear separation between the way we put world together and how it is put together is not accepted, the conceptual investigations can shed indirect light also on empirical questions. Wittgenstein’s position has been summed up by claiming that:

He accepted that philosophical problems and propositions are a priori in that they have their root not in reality but in the conceptual scheme we use in describing reality... he regarded this conceptual scheme as essentially embodied in language. He came to recognize... that language is not [an] abstract logical system... but rather a human practice and hence subject to historical change.

This, so it seems to me, captures also the spirit of Davidson’s work despite his Quinean naturalism. Quine held on to the idea that science is the highest path to truth, and this is a view rejected by Davidson, who turned philosophy’s focus more towards ordinary language.

1.2.3 Von Wright’s conception of philosophy

My claim is that von Wright’s understanding of what philosophy is and what a person is doing when she does philosophy are especially important in order to understand his views on the philosophy of mind. These kinds of questions are often ignored, but a discussion of them helps us to understand, partly, why a certain philosopher holds a certain position. Von

---

188 When describing his famous thought experiment of radical interpretation, Davidson (1980, 12) writes that “The approach to the problems of meaning, belief and desire that I have outlined is not... meant to throw any direct light on how we master our first concepts and our first language....” Exactly the same point is repeated ten years later (Davidson, 1990, 325), so on this issue Davidson’s views remained steady. He repeatedly emphasizes that the model of radical interpretation should not be seen as a suggestion for how we actually understand each other. See Davidson, 1973a, 1984, 1994, 1994b, 1995a, 1999d, 1999e, 1999f and 2000.


190 Glock, 2006, 295.
Wright’s views about the nature of philosophy changed during his career. I will not describe the subtleties of the different conceptions that he held. I focus on the view which he reached as a result of his intellectual journey, because this conception is the most relevant one for von Wright’s views in the philosophy of mind. Von Wright thought that it would be futile to search for a general definition of philosophy, but he thought that useful progress would have been made if we could appoint a task for philosophy which looked interesting in itself and which clearly differed from the tasks of special sciences\textsuperscript{191}.

Von Wright strongly emphasized that the question about the nature of philosophy or philosophizing was something which interested him.\textsuperscript{192} He traced this interest back to Wittgenstein by noting: “The most lasting and consequential way in which Wittgenstein has ‘influenced’ me concerns my view what philosophy is, of what I, as a philosopher, am doing.”\textsuperscript{193} Von Wright thought that this kind of reflection about the nature of philosophy is especially important at times when a change in paradigm can be detected. I believe that the challenge of science for philosophy, \textit{which is reality now}, is an example of a change in the way how to think about the nature of philosophy; we have seen how many modern philosophers of mind that the nature of philosophy \textit{is} changing. Although the naturalistic conception of philosophy is currently quite well accepted, surprisingly few philosophers pay attention to the question of whether there are features peculiar to philosophy which are missing from science or whether science and philosophy are altogether two very different kinds of activity. Few contemporary philosophers pay attention to the question of what is it that they are doing when they are doing philosophy. Von Wright is a refreshing exception to this consensus. Philosophers not considering the nature of philosophy pay insufficient attention also to the question of what it is that the philosophers who do \textit{not} share a naturalistic conception of philosophy are doing. Insensitivity to this question causes misinterpretations regarding their conclusions about non-naturalistic philosophy.

Von Wright, who did not ever arrive at a final conclusion concerning the question what philosophy is, claimed that certain principles were part of his understanding of the nature of philosophy and of the nature of philosophical problems\textsuperscript{194}. They are:

1) Philosophy is not one of the (natural) sciences.

2) Philosophy is an activity and not a doctrine.

\textsuperscript{191} See von Wright, 1947.
\textsuperscript{192} See von Wright, 1985a.
\textsuperscript{193} Von Wright, 2001, 179.
\textsuperscript{194} Von Wright, 1985a, 2001.
3) The purpose of philosophy is the clarification of meaning.
4) Philosophical problems originate from linguistic confusions.
5) Exposing the confusions makes the problems vanish.
6) Philosophical method is logical analysis.

These remarks bear an obvious resemblance to Wittgenstein’s views. It is not a coincidence that von Wright, who accepted a Wittgensteinian view about the nature of philosophy, thought that Wittgenstein’s most original contribution was perhaps his understanding about the nature of philosophy and its relation to science.

According to von Wright, philosophy attempts to understand and clarify meaning, whereas science aims at truth. Alien to this conception of philosophy is the idea that observations would be especially relevant to philosophy or that philosophy would be an experimental activity. By claiming that philosophy is in some sense an \textit{a priori} discipline, von Wright is clearly continuing the Wittgensteinian “philosophical project”. Although Wittgenstein was reluctant to form a school of philosophy and have followers continuing his work, he also thought that he had arrived at a new way of philosophizing which wasn’t merely a stage in the evolution of philosophy but rather the beginning of a new subject. The question about the importance and relevance of the “Wittgensteinian project” is currently in focus as the debate between Dennett / Searle and Hacker / Bennett shows. The conception of philosophy which Hacker and Bennett defend is in essential respects the conception to which also von Wright subscribes. The relevance of this conception is seriously challenged by Searle, Dennett, Churchland and others.

Some of von Wright’s principles have been discussed in connection with Wittgenstein. The claim that philosophical problems originate from linguistic confusions and that the problem vanishes when the confusions are exposed should be understood in the Wittgensteinian spirit. The same can be said about the views that philosophy is not one of the sciences and that philosophy is an activity and not a doctrine. I think that it is not entirely clear how the rest of von Wright’s defining principles should be understood. What exactly is “logical analysis”? How are “meanings clarified”? Instead of trying to find out what von Wright means by these expressions, one can consider the ways he actually did philosophy. One way to consider this is to see in detail how von Wright sees the nature of philosophical problems and how he tries to solve them. If philosophy should be understood as an \textit{activity}, how does von Wright pursue this activity? In his writings von Wright is often troubled by the question of what the source of \textit{philosophical motivation} is. He is trying to describe what is
characteristic of a situation in which philosophical motivation arises. I believe this is a relevant question for any contemporary philosopher to think about. This is especially relevant for a modern philosopher of mind, because there are various ways to study the mind and they are often in competition. What are those occasions, for example in the context of questions about the mind, when a philosopher instead of a psychologist or a neuroscientist is needed? What is the contribution of a philosopher? What could it be? What should it be? When does one encounter a question which requires philosophical treatment, that is, clarification of meaning in terms of conceptual analysis? How to tell whether such a question has been encountered? In a word, how to detect a philosophical problem? For the self-image of philosophy these are important questions, especially in a society where the highest measure of success (truth) is a clear result, which should be useful. If philosophical problems are not empirical problems, it is hopeless to think that they could be solved in terms of science. Likewise, a philosopher may be, and most likely is, useless when a scientific problem about the mind awaits a solution.

According to von Wright, what motivates a philosopher, and not perhaps a psychologist or a neuroscientist, is a confusion which a poorly set question creates. Such a question may contain a latent or an obvious inconsistency, which is possibly a result of the fact that words are being used in ambiguous ways or outside their usual fields of application. A philosophical question is thus a question which is a result of the misuse of language. How to solve this kind of problem? The source of philosophical motivation is a confusion caused by a poorly formulated question, but what is the aim of philosophy? The aim of philosophy is to calm the anxiety of thought by revealing the obscure conceptual structures which were the reason for the occurrence of a philosophical problem in the first place, and thus to provide a reason for the perturbation of thought. Philosophical clarification shows that there was no reason to be anxious, because the problem was wrongly stated. There was something wrong in the original way in which the question was formulated. Philosophical problems are therefore not solved, but they cease to be problems when our understanding about their nature increases. They cease to be problems when our thoughts regain peace. But it must be acknowledged that the problems are not solved through explanation. Von Wright claimed that it is a mistake to think that philosophical problems could be solved; the thought that these problems would have a final solution is “atrocious”.

---

195 Von Wright (1985) notes that philosophical problems cannot be solved but they may lose their importance, which is not to say that the same problems could not become important again.
196 Von Wright, 1979.
If philosophical problems are due to misuse of language and if the confusion which these problems create may be removed by analyzing language, how is this actually done? The question of how to approach a scientific problem is, in the context of a given science, often straightforward. The question of whether the problem has been solved is also often possible to answer; the criteria of success are relatively clear. But how to decide whether a philosopher has succeeded in clearing away a philosophical problem? Von Wright’s view is that a philosopher is, above all else, interested in concepts – and it is especially the concepts of everyday language which interest a philosopher. I want to emphasize that this is how von Wright describes his own interests as a philosopher. He does not claim that this is what philosophy should be. A philosopher in the von Wrightian mold feels that the use of everyday concepts is often unclear and therefore in need of systematization. What is of the utmost importance is that, as in Davidson’s case as well, it is concepts and not phenomena which are of interest in philosophy.197 Philosophy aims at the clarification of meanings and, contrary to science, is not interested in finding out how things are. Philosophy is an activity which clarifies thoughts in cases when language leads us astray. The way that the use of concepts can be systematized is by clarifying conceptual relations in a certain field of concepts. Different features of concepts can be revealed by relating concepts to each other. Both von Wright and Davidson think that a clarification or explication in terms of understanding is a correct method of philosophy. Von Wright’s “logical analysis” thus resembles a Davidsonian “clarification”.

The expression “field of concepts” is a metaphorical one, which is meant to emphasize that concepts cannot be analyzed in isolation. In order to say something illuminating about the concept of mind, for example, something about a great number of other concepts must be said. According to von Wright, the systematization or clarification of concepts means placing these concepts in order so that philosophical problems will not occur anymore. How this is to be done is, in my opinion, anything but clear. I would suggest that if philosophical problems are partly personal perplexities, then the question of whether a philosophical clarification has succeeded can be answered only from a subjective perspective. On the other hand, the work of a certain philosopher – for example Wittgenstein – can be seen as showing one way of how to place concepts in order. Von Wright notes that the activity which aims to put the meanings of concepts in a certain order could be called “descriptive metaphysics” or, as he also likes to call it, “conceptual phenomenology”. Incidentally, “descriptive metaphysics” is the term that,

197 Von Wright admits that the distinction between concepts and phenomena is not always easy to make.
according to a recent commentator, best describes Davidson’s approach to philosophical problems.\textsuperscript{198}

If concepts should be put in an order where philosophical problems do not occur, how is it possible to know what kind of order is to be searched for? Two suggestions were already given: On the one hand it could be claimed that the achieved order can be evaluated only subjectively; on the other hand, one could claim that works of different philosophers may show different examples of orders in which philosophical perplexities vanish. But these suggestions are very vague, and perhaps only “Wittgensteinians” would agree that Wittgenstein managed to show how to put concepts in an order where perplexities have completely disappeared. It could also be claimed that the question “what kind of order should one look for?” is itself a queer question. Why? Because in the case of concepts which cause philosophical problems, the question of what these concepts exactly mean is often an open question. A philosopher should start by asking “What do we mean by the concept of mind?”, and then the task of philosophical clarification begins.

Von Wright’s essential point is this: the use of a certain concept is unclear. A philosopher starts unraveling the vagueness by considering what we mean by the concept, by describing how it relates to other concepts. It is crucial that in doing so, \textit{a philosopher cannot appeal to the actual use of language because this use is imperfect}. A philosopher cannot appeal to the rules of how to use language because, in the case of problematic concepts or their specific use, \textit{such rules do not exist}. If there is anything to which a philosopher can appeal, \textit{it is her own intuition, which tells her what kind of rules will provide the clarity or systematicity that she is searching for}. This reminds one of Wittgenstein’s view, which von Wright quotes with approval: “Work on philosophy – like work in architecture in many respects – is really more work on oneself. On one’s own conception. On how one sees things. (And what one expects from them.)”.\textsuperscript{199} A philosopher’s work consists of explicating or interpreting certain conceptual intuitions of her own.\textsuperscript{200} This is an observation of crucial importance; it gives a certain specific view about the nature of the “philosophical method” and it is essential for understanding correctly von Wright’s philosophy of mind.\textsuperscript{201} Von Wright thought that a question is “logical” or “conceptual” – that is, \textit{philosophical} – if the

\textsuperscript{198} See Glock, 2003a.
\textsuperscript{199} Wittgenstein, 1980a, 16.
\textsuperscript{200} See, for example, von Wright 1985a, 1989a. This, once again, is the way von Wright sees the nature of philosophy. It is the way he sees his own work, but also the way in which he sees the nature of philosophy as a historical phenomenon.
\textsuperscript{201} For one view about the role of intuitions in philosophy, see Knobe and Nichols (eds.), 2008.
answer is at least *partly dependent on the answerer, on his suggestions and consideration*. A “creative component” contributes to the solving of conceptual problems.\(^{202}\)

Whereas a philosopher must appeal to her own intuitions, the actual use of language provides a *negative test*. Von Wright claims that a philosopher should not say anything which would go against the actual use of language. The language of the everyday is in the end the foundation on which we all must stand, and therefore the criticism that the philosopher raises against language must be exercised in the language of the everyday, and not by using some philosophical terminology invented for this purpose. A philosopher should not say anything which sounds strange or paradoxical. Philosophy should not be in conflict with common-sense.\(^{203}\) I think this is a very challenging requirement; it is difficult to achieve, and respecting the demand blindly will lead to bad philosophy. I also think that von Wright has nevertheless followed his own advice remarkably well; *In the Shadow of Descartes* is an excellent example of this. Just how strongly this kind of understanding of the nature of philosophy differs from the naturalistic conception of philosophy cannot be exaggerated. Especially, the claim that philosophy is the explication of the conceptual intuitions of the philosopher is something which would sound utterly wrongheaded from the perspective of naturalism.

*If* we agree with von Wright that – in the case of philosophical problems – how a certain claim should be understood is an open question, then the following observation follows. As von Wright notes, no matter in what way a philosophical problem is “solved”, certain assumptions must be rejected if others are accepted. A philosopher who defends eliminative materialism accepts certain assumptions, like the ideological assumption that results of science are privileged, and must therefore reject certain other “assumptions” like the existence of propositional attitudes. A philosopher who defends the common-sense conception of ourselves accepts, for example, that what is obvious has primacy and thereby rejects the view that science could eventually describe us as we “really are”. But what assumptions are accepted and what assumptions rejected – that is, in which ways the concepts are related to each other in a certain conceptual field – depends, partly, on the philosopher’s own *decision*. There is no way to determine which decision is the right one. The decisions that a philosopher makes are acceptable insofar as they provide a clear overall picture of the conceptual field to be clarified. Every attempt to clarify conceptual intuitions must accept certain concepts as basic, and other concepts are to be clarified in terms of these. But on what

\(^{202}\) Von Wright, 1947, 388.

\(^{203}\) Von Wright emphasizes that this is a very *personal* choice and that he learned this principle from Moore and Wittgenstein.
kind of foundation the conceptual clarification is to be built depends on the philosopher’s choice, and certain choices are such that they cannot be proven to be correct. In von Wright’s case, the view according to which ordinary language is a “transcendental” foundation of all philosophical investigations could be an example of this kind of choice, which cannot be proven to be correct but which nevertheless needs to be made.

My claim is that what is important in von Wright’s case is that the philosophical decisions he makes are strongly influenced by his views concerning much broader issues, such as the nature of humanity and society. It seems to me that in von Wright’s case, his philosophy and way of philosophizing was more of a personal agenda than in the case of many modern philosophers. We can assume that Wittgenstein strongly influenced this aspect of von Wright’s philosophy; von Wright often noted that it was especially Wittgenstein’s personality which had a strong effect on him. The exact nature of Wittgenstein’s influence on von Wright is a question which must be left to those who knew von Wright personally. It seems obvious to me that in von Wright’s case, his “philosophical” and “non-philosophical” views cannot be clearly separated. He wanted to defend a humanistic conception of human beings against the manipulation of humanity, and it is therefore no co-incidence that his writings in the philosophy of mind emphasize the irreducibility of the mental and the autonomy of psychology with respect to the natural sciences. Von Wright noted that the problem of humanism was something which preoccupied him throughout his mature life. In the center of this problem was “a search for an attitude to life and to one’s fellow humans”. Given von Wright’s will to defend humanism and his view that the answers that a philosopher provides affect his own judgments and the way that he relates to his fellow humans and society, it is a natural conclusion to think that von Wright’s arguments in the philosophy of mind are meant to have effects outside philosophy. Indeed, von Wright claims that insofar as the ideas of a philosopher influence the thinking of others, a philosopher influences the life of others. A philosopher is thus not only providing conceptual clarification, (s)he is also changing the nature of reality through changing the attitudes of humans. I believe the best way to understand von Wright’s mature conception of philosophy is that philosophy is a normative discipline; its results can and should have an effect on society.

Naturalists see philosophy as being a continuous effort with science. The purpose of philosophy is not the clarification of meanings but, together with science, to describe how things are. In order to understand von Wright’s philosophy in the right way, the most

---

204 For one such a view see Stoutland, 2006b.
205 Von Wright, 1989, 843.
important thing to realize is how alien von Wright’s conception of philosophy is to
naturalistic conception. Likewise, from von Wright’s perspective the kind of philosophy
which is done in the naturalistic spirit cannot address those questions which interested him.
By claiming that in philosophy Russell has defeated Wittgenstein, von Wright was acutely
aware that the naturalistic trend is currently dominant in philosophy. Philosophy has become
scientific and thereby distanced itself from the Wittgensteinian conception of philosophy. Von
Wright was very skeptical of this trend as well as towards the contemporary way of life in
general. In his intellectual autobiography, von Wright claims that it is important to distinguish
the peculiar activity of a philosopher from that of a logician, a mathematician, a natural or
social scientist. 206 The way in which the activity of a philosopher differs from the activities of
natural scientists has been discussed already. But we should acknowledge von Wright’s more
general belief that a society or culture where knowledge is emphasized at the cost of
understanding can lead to uncongenial results. An intellectual atmosphere which aims at
methodological monism may blur or ignore “the conceptual features peculiar to humanistic
study…” and result in the dehumanization of people, as von Wright notes. 207 Of these dangers
we should be aware. The conceptual confusions and problems resulting from the monistic
tendency of which von Wright had warned already in his Explanation and Understanding are
concrete now. Bennett’s and Hacker’s Philosophical Foundations of Neuroscience is an up-
to-date description of these conceptual problems. They would agree that it is the task of a
philosopher to “give a truthful general account of the conceptual peculiarities of this mode of
[humanistic] understanding”, as von Wright claims. 208 They would also agree, as their whole
project testifies, that the failure to recognize the distinct logical nature of this mode of
understanding has, in von Wright’s words “[…] fostered a tendency to look for
methodological ideals in a direction, viz., natural science, where they should not be
sought”. 209 With respect to methodological monism von Wright’s position is absolutely clear.
Already in Explanation and Understanding he subscribes to a form of conceptual dualism,
and the belief in a certain kind of irreducible dualism of nature and humanity was a guiding
principle guiding his work until the end. 210 This dualism is one of the most essential points of
von Wright’s work in the philosophy of mind. In one of his last writings, von Wright even

206 Von Wright, 1989a.
207 Von Wright, 1989, 844.
208 Von Wright, 1989, 834.
209 Von Wright, 1989a, 41.
210 It is telling that in the symposium conducted in his memory, von Wright was remembered by these words:
“[…] he argued that understanding in the human and social sciences involves intentionality and practical
reasoning in a manner that distinguishes them from the causal explanatory patterns of the natural sciences.”
(Niiniluoto, 2005, 10)
went so far as to describe his unquestioned basis as “Cartesian”. By the use of this label von Wright wanted to emphasize that in his view there are “two kinds of phenomena which are (conceptually) sharply distinguished and irreducible the one to the other.” This Cartesian standpoint von Wright describes as a form of non-reductive dualism.

The position of methodological dualism stands opposed to a view which emphasizes the role of knowledge as the highest cognitive activity of human beings. If we consider the current intellectual atmosphere in Western societies, I believe we can conclude that von Wright’s worry – according to which hypertrophy of knowledge may happen at the cost of understanding – is legitimate and perhaps currently more relevant than ever. In von Wright’s opinion, a good intellectual education or intellectual culture is one in which the two cognitive aspects of humanity, knowledge and understanding, are in balance. In his view, there exists currently an imbalance because although we know more than ever, we do not always understand correctly what it is that we know. Von Wright thought that this was one sign of the overall sickness of modern society. We focus on increasing our knowledge without realizing that real spiritual progress comes only when knowledge is accompanied by understanding. I would like to think that the need for philosophy in addition to science is connected precisely to the need to understand the phenomena of our time. Those thinking that philosophy competes with science are apt to conclude that the relevance of philosophy is decreasing as the amount of scientific knowledge is accumulating. Philosophy’s progress pales when compared to the progress of science. Quantitatively, science rises above philosophy, but this is because, as von Wright notes, there is no progress but “regeneration” in philosophy. I think Hacker captures very well the general view to which von Wright would subscribe: “[...] the form of science is progress, the structure of science is hierarchical, and advances in science are built upon prior achievements. Not so in the case of philosophy, which is ‘flat’. Achievement in philosophy is the attainment of understanding, not the acquisition of fresh knowledge.”

With the increase in knowledge comes an increased need to understand what we know. As we will see, von Wright’s work in the philosophy of mind is an attempt to better understand, for example, the claims of neurophilosophers. In order to understand what it is

211 Von Wright, UPe, 1.
212 This expression is rarely used in the contemporary philosophy of mind. It is unclear whether there are any philosophers who would describe their position in these terms. The content of von Wright’s position will be described later in sections 2.2, 2.2.1 and 2.6.
213 For von Wright’s views on this subject, see his interview with Kuusela, Lahtinen and Pihlström, 1998.
214 See von Wright, 1981.
that is being claimed, a conceptual clarification is often required. Von Wright’s writings in the philosophy of mind are not hypotheses competing with the claims made by “scientific philosophers of mind”. His approach goes “deeper”, because the purpose is to understand to what extent these claims make sense. This is a straightforward consequence of von Wright’s understanding concerning what a philosopher should do.

### 1.3 A methodological caveat

I have suggested that von Wright and Davidson could resist the naturalist challenge for philosophy given their Wittgensteinian understanding about the nature of philosophy. I think there is, however, an inconsistency between their conception of philosophy and the way they actually practice philosophy or put forward philosophical theses. A “true Wittgensteinian” like Hacker would challenge the assumptions of modern philosophy of mind in a quite different way than Davidson and von Wright.

The observations that must be made for the coming chapters are these: 1) Davidson and von Wright could have used their understanding about the nature of philosophy more effectively against reductive versions of naturalism, thereby emphasizing the need for a certain kind of philosophy and for certain kinds of philosophical arguments. They could have emphasized more strongly that philosophy differs from science and could have used this distinction to justify their philosophical theses. 2) Davidson and von Wright only rarely utilize their conception of philosophy and are actually often playing “the intellectual game” of the reductive naturalists. 3) Since I want to analyze the philosophies of von Wright and Davidson, I am not attempting to use the Wittgensteinian conception in order to challenge the scientific or scientistic philosophy of mind. My purpose is to show both situations: those on the one hand where von Wright and Davidson are not true to their conception of philosophy and seem to be playing the game of scientific philosophers, as well as those on the other, where they do seem to challenge contemporary philosophy based on their Wittgensteinian conception.

In this work I am willing to accept many of the assumptions of the reductive naturalists. The reason for this is twofold; first, I want to show that reductive naturalism and its physicalistic ontology face serious problems even if we are willing to grant certain basic premises of the position instead of dismissing the claims simply as nonsense, like Hacker and many other neo-Wittgensteinians do. I am therefore willing to let the reductive naturalists choose the place where the fight over the human mind will be fought, and I am also letting them bring many of their own weapons with them. Another, more straightforward critical
strategy, would be to change the rules of the game as, for example, Hacker does. By following this strategy, one unfortunately also marginalizes one’s own position and thereby makes it less persuasive.

I believe that the confusions besetting the modern philosophy of mind could be better tackled by the use of Wittgensteinian insights than is currently being done. This challenge must, however, be the topic of future work. The other reason why I am willing to grant many of the assumptions of my opponents is the fact that this is what Davidson and von Wright did as well. Despite their conception of philosophy, their positions were closer to a kind of naturalistic philosophy of mind than to the kind of position argued by Hacker. My purpose is to explore how Davidson and von Wright, while being critical of naturalism, manage to respond to the naturalistic challenge in the context of naturalism.

1.4 Conclusions of chapter one

In this chapter I have tentatively suggested that science is becoming an increasing threat for the philosophy of mind. I have clarified the intellectual landscape in which the contemporary philosophy of mind operates. The relevance of the kind of philosophy that Davidson and von Wright can be seen as practicing is challenged in this intellectual atmosphere, which encourages scientific philosophy. This raises the question of how philosophy in general, and von Wright’s and Davidson’s positions in particular, can resist empirical pressures.

In section 1.1 and 1.1.1, the possible consequences of a naturalistic conception were considered and the general problems of eliminative materialism, which is a paradigm example of scientific philosophy of mind, were discussed. In section 1.2, I made it clear how Davidson and von Wright could answer the challenge posed by science. This was done by considering what, in their view, the purpose of philosophy could be and by noting that science cannot fulfill this purpose. In clarifying von Wright’s and Davidson’s conceptions of philosophy, I have emphasized the similarities between their views and suggested that Wittgenstein’s understanding of the nature of philosophy has influenced both von Wright and Davidson. Their positions in the philosophy of mind cannot be properly evaluated if their conceptions of philosophy are not taken into account.
Chapter two: Non-reductive physicalism and its warrantability

In this chapter I consider the following question: How should we understand the physicalism of ‘non-reductive physicalism’, and is such physicalism warranted? This question has been one of the most pressing questions in the modern philosophy of mind. My purpose is to discuss whether non-reductive physicalism is a satisfying answer to the classical mind–body problem. Many modern philosophers claim that the mind–body problem has been solved; mind is a physical thing, or at least a physical phenomenon. According to John Searle:

> The famous mind–body problem... has a simple solution. This solution has been available to any educated person since serious work began on the brain nearly a century ago, and, in a sense, we all know it to be true. Here it is: mental phenomena are caused by neurophysiological processes in the brain and are themselves features of the brain.

This is the claim of naturalism. Science has solved the mind–body problem by showing that mental phenomena are features of the brain. “We all know it to be true”, as Searle claims. Yet there are philosophers who think that this conclusion of naturalism is rushed and oversimplified. Thomas Nagel claims that: “We do not at present have even the outline of an adequate theory of the place of the mind in the natural order.... Our knowledge is entirely empirical and ad hoc, not theoretical” and suggests that since “no one has a plausible answer to the mind–body problem, all we can really do is to try to develop various alternatives one of which may prove in the long run to be an ancestor of a credible solution.”

Colin Mcginn notes: “We have been trying for a long time to solve the mind–body problem. It has stubbornly resisted our best efforts. The mystery persists. I think the time has come to admit candidly that we cannot resolve the mystery.” There is thus disagreement among the most active philosophers of mind with respect to question whether or not the mind–body has been solved, or even whether it can be solved. Physicalism is one answer to the mind body problem, but many see its status as being that of an empirical hypothesis. My interest in considering the mind–body theories of Davidson and von Wright is to clarify whether they have plausible philosophical, i.e. in some sense non-empirical, arguments for physicalism. Because the problem persists, I believe that one should be open to different views as Nagel suggests. What I want to find out is whether von Wright and

---

216 Searle, 1994, 277.
217 Nagel, 1994, 63.
220 See, for example, Place, 1960 and Foss, 1987.
Davidson present something philosophically unique and important for the contemporary mind–body debate, and how they would answer the main question about the status of non-reductive physicalism. In doing this, I will clarify several exegetical issues by showing the similarities between the positions of von Wright and Davidson. This exercise demonstrates the current relevance of these philosophers by showing how they can contribute to the contemporary discussion. The sub-questions that will be considered are the following: What is the content of the physicalism to which Davidson and von Wright subscribe? What reasons do they give for this kind of physicalism? How do they address the problems that a non-reductive physicalism faces? Would it be more correct to describe their non-reductive positions as versions of non-physicalism? The sub-questions that will be considered are the following: What is the content of the physicalism to which Davidson and von Wright subscribe? What reasons do they give for this kind of physicalism? How do they address the problems that a non-reductive physicalism faces? Would it be more correct to describe their non-reductive positions as versions of non-physicalism? The emphasis will be on Davidson’s view, given the influence of anomalous monism (AM) for the contemporary philosophy of mind and given the fact that he wrote more on the subject that von Wright.

Most contemporary philosophers are physicalists. What a commitment to physicalism requires or implies, however, is a matter of current disagreement. How the position of physicalism should be exactly defined is also a question under debate. The amount of literature discussing these very questions is a testament to how unsettled even the basic questions about the nature of contemporary physicalism are. One common way to describe physicalism is to say that this position claims that there exists only one substance and that the nature of this substance will be described, at least in principle, by the most general science that studies the nature of reality. This science is, by definition, physics. Everything that can be thought of as being constituted of something is somehow constituted of the basic entities that fundamental physics describes, and everything is governed by the fundamental laws of physics. Philosophers now widely accept that immaterial souls do not exist and nothing lies outside the spatio-temporal physical reality that physics studies. Seen from a historical perspective, this consensus about the truth of physicalism is a very recent (namely a 20th century) view. In the history of philosophy, substance dualism, idealism or some other non-materialist position has often been seen as a real alternative by many first-class philosophers.

What has convinced the majority of philosophers of the truth of physicalism? Fodor has noted that one reason to believe that physicalism is true is the fact that the alternatives are even worse. This humorous comment invokes a partial truth, but it is obviously not a good reason for a critical philosopher to think that physicalism is de facto true. Why is it the case that all non-physicalistic alternatives seem implausible nowadays? The currently existing

---

221 The rejection-rate of the notion of ‘soul’ seems to be higher among philosophers than among the general population. The impact of science on philosophy is perhaps a major reason for this.
ontological consensus in the philosophy of mind is, at least partly, a straight result of the program of naturalism which has been dominant in analytical philosophy since the 1950s. As has been discussed in chapter one, the main claim of naturalism is that philosophy should be continuous with science. Scientific knowledge about the nature of reality, the “scientific world view”, suggests that the whole of reality is physical. Many philosophers think that there is no real evidence for dualism. By this they mean that there is no scientific evidence for this position. In the context of naturalism the interest in philosophical arguments for dualism has decreased. There is no point in defending or even exploring a dualistic position, because it would be at odds with the results of science. Because there allegedly is scientific evidence for physicalism, there is not much interest in defending physicalism by philosophical arguments either. Physicalism is a position which nowadays is often taken for granted; it is a starting point from which considerations about the problems of the mind proceed.

As far as ontology (the question of what there is) is concerned, dualism has been rejected and physicalism has been accepted. What does the acceptance of physicalism actually tell us about the nature of the mind? For many contemporary philosophers the mind–body problem, as traditionally understood, has lost its importance and appeal. Neurophilosophers are a paradigm example of this, because their work is done completely in the context of physicalism. Whereas in the history of philosophy the mind–body problem has been the problem of whether mind and matter are identical, whether mind is matter or vice versa, the current problem is seen as one of how the physical mind fits to the physical world. The question is not whether or not mental phenomena are physical phenomena. In this sense, the ontological mind–body dichotomy has been erased. The resulting question is what kind of physical phenomena mental phenomena are. This, according to many naturalists, is best understood as an empirical question, which will be solved in terms of the sciences studying the brain.

But what does it mean to say that physicalism is true with respect to the mind? How should the content of physicalism be understood? What follows from the acceptance of physicalism? How well does a physicalistic position answer the question of what our nature as thinkers and agents is? Does the acceptance of physicalism mean something more than accepting the “mere” claim that the mind is identical with the brain and therefore beliefs, desires, hopes and wishes are states of the brain? With respect to these questions, two versions of physicalism – reductive and non-reductive – stand far apart, even though both are committed to the view that “everything is physical”. But, or so I would like to think, the claim that everything is physical is empty given that mind has not been reduced to matter. What
does it mean to say that the mind is a physical entity? It is not known how brains produce
mental phenomena, if indeed they do. The relation between brains and mental phenomena, if
there is such, is unknown. The occurrences of specific mental phenomena cannot be clearly
located. We cannot explain how mental phenomena can cause bodily movements. How the
idea of free will fits into a physical world view is a mystery. The physical status of subjective,
qualitative aspects of experiences remains puzzling. All these questions are more or less
unresolved; yet the consensus view is that mind is a physical entity, as if this claim could be
understood in any clear sense in the absence of answers to the previous questions. “Mind is
nothing but the brain”, goes the famous slogan. What does this claim actually tell us about the
mind or about our mental lives? Does this slogan tell us anything about the most important
aspect of ourselves, of our mental lives? Does it offer anything more than an abstract
suggestion that we are part of nature? But what does this mean? We, as persons, feel love,
sadness, joy, jealousy, for example. We plan things, have hopes, act freely according to our
wishes. Sometimes we act against our better judgments or do things that we should not do.
We wonder what the “meaning of life” is and spend endless hours thinking about how we
should live our lives, often wondering what the right choice in a given situation is. Does the
physicalistic answer give us any advice on how to think about these essential aspects?

These questions show that the position of physicalism, although a widely accepted
view in the current philosophy of mind, is full of unresolved questions. What I will consider
here is whether Davidson and von Wright give convincing arguments for physicalism. As will
become clear, these philosophers share the view that everything there is, is physical. In this
sense they are part of the physicalistic movement of modern philosophy, although it would be
misleading to say that they belong to the naturalistic movement in any substantial sense.

Davidson and von Wright accept physicalism insofar as it is understood as a view opposed to
substance dualism, i.e. a modern view according to which nothing exists except “physical
particles”, their constitutions and forces, as recognized by physics. This position can be called
O-physicalism. But the term “physicalism” is also often associated with a view that an
adequate account of mental or other “problematic” phenomena must either be an explanatory
theory from the natural sciences or be somehow vindicated by such a theory. This position

---

222 It has to be admitted that the situation is not as straightforward and pessimistic as the claims above may
suggest. Of course scientists already know or claim to know many things about the relation between mind and
brain. Nevertheless, it is widely accepted among many philosophers that consciousness is still a very deep
mystery and the problem of free will is as pressing as ever.

223 As noted in chapter one, both Davidson and von Wright do express a commitment to naturalism by denying
the existence of “super-natural” entities. This form of naturalism is so weak that it needs to be clearly separated
from the more substantial forms of contemporary naturalism that von Wright and Davidson reject.

224 O is for ontological.
can be called M-physicalism.\textsuperscript{225} An essential part of M-physicalism, together with the view that mental phenomena need scientific vindication, is that all problems about the mental can ultimately be explained in physical terms. Contemporary naturalism accepts the truth of O- and M-physicalism, and claims that the problem of consciousness, the problem of mental causation or the problem of free will can be explained in scientific, physical terms. On this kind of naturalism is based the view that physicalism should be thoroughgoing, thus encompassing ontology as well as methodology. According to this variety of naturalism, only truths entailed by the truths of the natural sciences should be believed, and what is inconsistent with these latter truths should be disbelieved. M-physicalism is something which von Wright and Davidson, by their rejection of this contemporary naturalistic view of philosophy, are willing to reject.

2.1 One answer to the mind–body problem: Davidson’s anomalous monism

In the following three sections, the question of whether Davidson’s theory of mind, anomalous monism, is a convincing physicalistic answer to the mind–body problem will be considered. Most critics have concluded that it is not. A general problem about Davidson’s position is that it is not clear whether it should be seen as a form of physicalism at all; if not, then, despite its other merits, it surely cannot be a physicalistic answer to the mind–body problem, whether in its traditional or in its or contemporary form. From a critical point of view, AM may thus fail in two ways in being a physicalistic answer to the mind–body problem. First, it may fail by being an inadequate answer against dualism. Second, supposing that it is an adequate answer against dualism, it may fail as not being an adequate form of M-physicalism. The latter failure would imply AM’s inadequacy as a naturalistic theory of mind. Davidson’ position would thus be hopelessly unscientific from the viewpoint of naturalism.

Before the relevance of Davidson’s views about the mind–body problem can be evaluated, they need to be made clear. Given that Davidson’s position has been described at least as a form of “ontological monism of events (respectively things) and a psychophysical dualism of properties that corresponds epistemologically to a dualism of concepts- and descriptions”\textsuperscript{226}, “emergent materialism”\textsuperscript{227}, “neutral monism”, “transcendental physicalism”\textsuperscript{225},

\textsuperscript{225} M is for methodological.
\textsuperscript{226} Rogler and Preyer, 2004.
\textsuperscript{227} Niiniluoto, 1990.
“property dualism” and “eliminative materialism”, there is a clear exegetical question to be solved. All the labels cannot be correct although, as will become clear, different interpretations of Davidson’s position are justified. In section 2.1.1 the general background of AM will be presented and the exegetical issues considered. In sections 2.1.2 and 2.1.3, the exact nature of the monism of AM will be discussed from the assumed perspective that AM is meant to be a physicalistic answer to the mind–body problem. This assumption will be justified in the next section.

2.1.1 The place of AM in contemporary philosophy of mind

With respect to philosophers who are no longer alive, there are two important questions to be considered. First, what, if any, was the relevance of the work of that philosopher when the philosopher was active? Second, what, if any, is the relevance of the work today? With respect to Davidson, who produced philosophical insights for decades, a widespread opinion seems to be that the ideas were original and relevant when they were put forward, but time has treated them unkindly. Here I will consider these two questions and aim for a conclusion about the relevance of AM based on the answers to them. Whereas some could argue that the first question is not interesting, I suggest that it is important to acknowledge how a position like AM has shaped, if at all, modern philosophy of mind. Moreover, following Nagel’s advice, I suggest that one should have interest in various kinds of mind–body positions. The materialist consensus should not entirely obliterate the interest of peculiar positions in the philosophy of mind. In the spirit of chapter one, I suggest that these positions may teach us more than most of the materialist positions.

There are not many contemporary philosophers of mind who succeed in constructing an original and distinctive position. Davidson is one exception because AM, a well known and an original view, is associated entirely with him. Davidson really came up with his own theory of mind, which he saw as an alternative to nomological monism, like physicalism, to nomological dualism, like parallelism, interactionism and epiphenomenalism, and to anomalous dualism, like Cartesianism. In the contemporary discussion, AM is usually interpreted as a form of non-reductive, token physicalism. It is a version of physicalism which is meant to be an alternative to other physicalistic positions like type-physicalism and eliminative materialism. Davidson is thus often interpreted as the originator of non-reductive
monism. This is a noteworthy achievement given the importance of this position today. The historical origins of AM can be detected in Spinoza and Kant. Davidson himself notes that, before he presented the view, a similar kind of position is recognized as a possibility by Herbert Feigl, Sydney Shoemaker and Charles Taylor, accepted by Thomas Nagel and endorsed by Peter Strawson. There are thus certain historical influences, but the formulation of the position and the argument structure is Davidson’s own. AM, although found to be an interesting position by some, does not have many contemporary followers. The most notable philosopher who came to endorse AM is in fact Quine. Rorty is also influenced by Davidson’s view, and to a lesser extent the same seems to be true of Nagel. Some contemporary philosophers show positive interest in AM although they do not fully subscribe to it. The general view of the status of the status of AM is, however, critical – and whereas there are, for example eliminative materialists discussing philosophical problems of the mind, there are no “anomalous monists” participating in the discussion.

In order to evaluate AM correctly, the situation in the philosophy of mind at the time when Davidson presented his position needs to be recognized. Although the mind–body problem is currently formulated in the framework of O-physicalism, there was a time, not so long ago, when the mind–body problem, as traditionally understood, was precisely one of the main issues on which the truth of ontological physicalism hinged. The sudden swing to the truth of physicalism is a surprisingly recent view. Ullin T. Place, one of the originators of type–type identity theory, noted in 1989 that:

[...] whereas in 1956 every philosopher you met was quite convinced that whatever answer to the mind–body problem, if there is one, is true, materialism must be false, today it is almost as difficult, at least in the English-speaking world, to find a philosopher who is prepared to defend any other position.

I would say that this sudden change in only a fifty-year timeframe is remarkable given the fact that what is at stake is an all-encompassing metaphysical position. Those who are not convinced that the truth of physicalism is an established fact of course still think that the mind–body problem is something which creates a serious obstacle to a physicalistic world view. According to the critical voices, the problem is not merely one of finding a place for the mind in the physical world. Given the unresolved mind–body problem, the very coherence of the physicalistic world view is problematic. I think that the current critical statements

228 See, for example, Macdonald, 1989.
230 Place, 1989, 19.
against physicalism show that the basic questions about the coherence of this position are still, pace reductive and eliminative naturalists, in need of serious consideration.\textsuperscript{232}

When the physicalistic consensus was not as firm as it is today, it was unclear how physicalism as a general view or as a view in the philosophy of mind could be argued for, shown to be true and defended. As Place notes, in the 1950s most philosophers thought that materialism was not an adequate answer to the mind–body problem, and an argument was required to change this opinion in favor of materialism. An interesting question, therefore, is what the factors that convinced so many philosophers about the truth of physicalism were. Jaegwon Kim, Tyler Burge and others have argued that certain scientific achievements were important for philosophers to start believing in this all encompassing metaphysical thesis.\textsuperscript{233} This is, of course, partly a sociological story about the progress of philosophy in the 20\textsuperscript{th} century, the details of which remain beyond the scope of this work. However, the scientific evidence started to point towards the conclusion that no matter which part of reality is studied, the entities and phenomena that are encountered are ultimately physical. Burge suggests that success in biochemical research as well as in non-human animal neurophysiology during the 1950s could be among such important scientific achievements. It seems to me that the increase in our scientific knowledge about the nature of reality created an atmosphere in which there was room – and also a need – for physicalistic theories in the philosophy of mind. David Papineau argues that physicalism came to prominence during the 1950s and 1960s in the philosophy of mind because there was enough evidence for philosophers to start believing in the completeness of physics.\textsuperscript{234} This, together with the kind of physiological evidence that Kim and Burge emphasize, convinced many philosophers. A metaphysical thesis was thus a result of scientific progress and not a result of philosophical reasoning, as many of the metaphysical conclusions in the history of philosophy had hitherto been. The consensus view concerned not only the general structure of the world but contained a specific view about the mind too, namely the view that since O-physicalism is true, the traditional mind–body problem will eventually be settled empirically through science.

But can the truth of ontological physicalism be defended simply by declaring that everything there is, is what physics or some other natural science describes there to be? Somebody’s awareness of herself, of the fact that she is conscious, that she has feelings etc. is certainly not a scientific truth. Such facts are not known to an individual because of science. It

\textsuperscript{233} See Kim, 1998 and Burge, 1992.
\textsuperscript{234} Papineau, 2001.
is not a same kind of claim as, for example, the claim that water is H2O. Whereas the latter claim can be explicated in scientific terms, science cannot explain why I am aware of being conscious or what my being conscious consists in, and there are no scientific breakthroughs which were required for us to become aware of our own consciousness, whereas a scientific breakthrough was required for us to believe that water is H2O. Through self-observation, one may feel that the subjective point of view of an agent is so unique to the agent in question that the grasping of that point of view is forever beyond objective science. The problem of the point of view of an agent, the problem of what being somebody actually means, or how it should be understood, raises deep existential questions about the meaning of life which cannot be discussed here.

If trust in scientific evidence is all that is required to believe that O-physicalism with respect to mind is unproblematic, why was the consensus not in place much earlier? For example, Darwin’s Origin of Species published in 1859 was for many a convincing proof that humanity is part of the nature in exactly the same sense as other animals are. Many think that the human soul was banished when it was shown that our ancestors were non-human primates. Yet, philosophers did not turn into physicalists after reading Darwin. But if the completeness of physics, an empirical claim which is a debated issue, is now thought to demonstrate the falsity of ontological dualism, why wasn’t the evidence that humans are sophisticated animals already enough? Without trying to solve, beyond the conclusions of Papineau, Kim and others, the puzzle of why philosophers started to believe in physicalism, I would like to suggest that the general prevailing attitude in a society at large is also a factor influencing also the direction of philosophical theories. The latter part of 20th century was a time when trust in scientific progress was high, and the same intellectual atmosphere still continues, and is in fact stronger than ever. Rorty, who often has the ability to see philosophical problems in a wider perspective, noted already in 1972 that:

> [...] the reason Cartesian dualism is so unpopular nowadays is not because of any application of the powerful methods of modern analytical philosophy, but simply because we keep reading in Life and The Scientific American about cerebral localization, the production of any desired emotion, thought, or sense impression by the insertion of electrodes and the like.\(^{235}\)

If Rorty is correct, the exorcism of dualism had nothing to do with philosophical arguments. Whether the scientific mood and therefore also the nature of philosophical theories about the mind are going to change during this century remains to be seen. Pessimism about science

\(^{235}\) Rorty, 1972, 219.
could also change the way how the autonomy of the mind with respect to scientific results is seen. A possible revival of religions could also cast doubts upon the prevailing all-encompassing materialism.

As an observation about the history of philosophy of mind, it can be noted that once the general metaphysical assumption that everything is physical was in place, different versions of philosophical identity theories of the mind became quickly popular. Identity theories seemed reasonable given our increased knowledge about the nature of non-mental reality. In the 1950s and 1960s the prospects of type–type identity theory were widely discussed. According to this view, whenever a person has, for example, pain she is in a certain neural state, so that she is in this state whenever she is in pain. It can thus be asserted that “being in pain = being in a neural state N”. Another position which had many supporters was eliminative materialism. As has been discussed already, this view suggests that eventually the need to establish identities between brain states and mental states vanishes because mental terms do not refer to anything that is real. Both of these approaches were strongly influenced by the view that philosophical results should be continuous with scientific ones. Kim recalls the situation of that time. The approach of type–type identity theorists:

[...] sounded refreshingly bold and tough-minded, and seemed in tune with the optimistic scientific temper of the times. It was an intriguing and exciting idea that mental events could just be brain processes, and that scientific research could show this, just as science showed us that light was electromagnetic radiation, and that genes were DNA molecules.236

Once again, we should be astonished by the fact that only fifty years ago the view that mental events are brain processes was “exciting and intriguing”, whereas nowadays many philosophers, like Searle, find this view dull and uninteresting and a truism at best.

Identity theories were thought to describe how the mind fits the scientific picture of the world which philosophers had already created for themselves. J.J.C. Smart, one main figure behind the type–type identity theory, notes that he sees it as necessary to defend a physicalistic world view for reasons of “scientific plausibility”.237 As a curiosity, it can be noted that von Wright wanted to argue against dualism on the same grounds, although he wasn’t especially interested in defending a physicalistic world view. Smart’s approach focuses on showing that mental states, the most difficult category for a physicalist of that time, can be incorporated into the broad physicalistic picture. His classic paper, “Sensations and Brain Processes”, begins with a very honest and sincere declaration that the author cannot

---

237 Smart, 1963, 657.
believe that consciousness would be something that could stay outside the physicalistic picture of the world. We are told: “That everything should be explicable in terms of physics… except the occurrence of sensations seems to me to be frankly unbelievable.” Smart admits that this is largely a “confession of faith” and goes on to say that a philosophical argument which would compel us to believe in “nomological danglers” would, in his opinion, most likely be flawed. But back in Smart’s times “scientific plausibility” was the only reason why dualism seemed implausible!

Were physicalistic identity theories arguments for physicalism or consequences of a physicalistic world view? Was one motive behind them to provide an additional argument against substance dualism and in favor of physicalism? Philosophers tried to provide reasons why physicalism with respect to mind should be accepted, but whether these theories were really required to defend physicalism as a general thesis is unclear given that a belief to general monistic position was reached on independent grounds, that is, by considering the empirical evidence and reading The Scientific American. This being said, it should be acknowledged that since a satisfying physicalistic theory of the mind remains to be discovered still, a convincing philosophical argument for physicalism would certainly be a good thing for a physicalist. No wonder philosophers keep writing books in which the physicalistic position is explicated and defended. Illustrative recent examples are John Bickle’s Philosophy and Neuroscience, a Ruthlessly Reductive Account, Andrew Melnyk’s A Physicalist Manifesto: Thoroughly Modern Materialism and Thomas Polger’s Natural Minds.

Although the motivation to argue in favor of type–type identity theories increased as a result of scientific discoveries, many identity theories were nevertheless “philosophical” in the sense that they were not based on empirical discoveries about the relationship between mental and physical phenomena. They were proposals and they were speculative. These “theories” were not, and did not become, scientific theories which are supported by empirical evidence from scientific inquiry. Type–type theories held that mental phenomena could be reduced to physical phenomena, and that a successful reduction would show and thus

---

238 Smart, 1959, 142. Similar kinds of remarks have been made to suggest that it is incomprehensible how brain could produce mental phenomena. An advocate of this view could say: “That mental phenomena could be explicable in terms of physics seems to me frankly unbelievable. This is largely a confession of faith, but an argument which would suggest the opposite would most likely be flawed.” Thomas Huxley (1866, 193), for example, once wrote: “But what consciousness is, we know not; and how it is that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the Djinn when Aladdin rubbed his lamp in the story, or any other ultimate fact of nature.” Huxley who was a strict follower of Darwin was nevertheless amazed when confronted with the mystery of consciousness; such amazement is still valid today.

“guarantee” the identity between the two. What was central to this view was the thought that reduction between the mental and the physical would proceed through laws linking the two conceptual or ontological domains. Whatever differences there were in the details between versions of type–type identity theories, all were forms of nomological monism for which Davidson offered AM as an alternative.

In 1970 Davidson offered an important contribution to the discussion of the time when he introduced his own theory of mind. It is a view which is meant to reconcile three principles which seem to be inconsistent, and yet, or so Davidson argued, all are true. They are:

1) The Principle of Causal Interaction: At least some mental events interact causally with physical events.
2) The Principle of the Nomological Character of Causality: Events related as cause and effect fall under strict deterministic laws.
3) The Anomalism of the Mental: There are no strict deterministic laws on the basis of which mental events can be predicted and explained.\(^{240}\)

A realization that 1–3 can be true together opens the possibility for AM. What was the nature of the relation which Davidson took to exist between the mental and the physical? His view is that there is only one substance, thus the monism of his position. A fairly standard interpretation is that in this respect, the view which Davidson put forward was meant to be an argument in favor of physicalism. AM is a form of O-physicalism. However, the other part of the view, namely the anomalousness, suggested that there are no laws linking the mental and the physical domains. So, although Davidson’s view can be seen as a version of an identity theory, it differs from the standard versions in an important and in a radical sense. What are the differences?

First of all, it seems to me that the overall motivation behind AM is quite different than the spirit which motivated the type–type identity theories of the 1950s. These theories were “tough-minded and in tune with the optimistic scientific temper of the time”.\(^{241}\) Kim has noted: “The mental autonomy that Davidson wants is not the kind of autonomy that philosophers like Jerry Fodor have sought for psychology…. Rather, it is the autonomy of agency and the will, of the kind that Kant famously sought”.\(^{242}\) The speculative type–type

---

\(^{240}\) See Davidson, 1970.
\(^{241}\) Kim, 1998, 2.
\(^{242}\) Kim, 2003, 125.
identity theories were the science-inspired philosophical views about the mind of their time, and Fodor’s speculative view, together with many naturalistic theories, are examples of science-inspired philosophical views of our time. Whereas there is nothing distinctively unscientific in AM, scientific evidence plays no part in Davidson’s version of the identity theory and it does not motivate the philosophical position. The reason for coming up with such a position was philosophical: a need to answer the Kantian challenge. As is well known, the challenge is: if physical determinism holds for all of nature, how is freedom possible? A current formulation could be: since determinism holds for all of nature, how is freedom possible? This problem will be considered in chapter four. I believe that an adequate answer to this question requires the development of Davidson’s and von Wright’s insights and a rejection of their version of physicalism.

The fact that AM has a deep philosophical aim distinguishes it from most contemporary materialistic theories of the mind and is part of its “attractiveness”. Davidson was not satisfied with the standard view in analytical philosophy according to which there is no conflict between determinism and free will. The claim was, and is even more today, that they are compatible. According to Davidson, this answer is radically incomplete, although he does not give clear reasons for his critical view. Davidson’s purpose nevertheless is to unite the mind–body problem with the question of the possible compatibility of free will and determinism. One can here recall what Wilfrid Sellars famously claimed: “The aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term.” If these words are believed, then – or so I think – AM is a better philosophical view about the mind than most of its rivals. It serves philosophical aims better than scientific “philosophical” models of the mind, and is therefore an interesting philosophical position. AM is not merely a technical view about the relationship between mind and body, but part of Davidson’s general picture of the human condition. If we require, as I think we should, that a philosophical theory of mind should say something about human freedom, then Davidson’s theory ranks higher on the scale than even a detailed view about the relation between mind and matter that nevertheless remains silent about the problem of freedom.

243 I described in chapter one how, according to Davidson, “philosophy is the only hope” for those who are troubled by the conflict between determinism and human freedom. It is thus not surprising that he constructed a philosophical argument for the identity theory which was also meant to address the conflict.
244 As was made clear in the first chapter, my claim is that this applies generally to Davidson’s philosophy. It is philosophy with notably philosophical aims.
245 For a recent defense of this, see Dennett, 2003.
246 Sellars, 1962, 37.
Second, a more technical difference between AM and the identity theories of the 1950s and 1960s is Davidson’s claim according to which there are no laws linking the mental and the physical realms. When the view was first offered its novelty was that the argument for physicalism was given without the claim that there have to be laws linking the two domains. In fact, although “Mental Events” from 1970 is the classic paper to which references are always made in the subsequent literature, Davidson had already noted in 1963 that his position seemed to be identical with the views of Herbert Feigl and Smart if it were not for his claim about the anomalism of the mental. Feigl had offered his view in 1958 and Smart in 1959. Their purpose was to defend physicalism. In my opinion, Davidson’s reference to these type–type theorists can be interpreted as suggesting that his theory was also meant to offer support for physicalism. Although AM has philosophical aims, perhaps one motivation behind it, as in the case of Smart, was to provide support for physicalism for reasons of “scientific plausibility”. But the distinctive aspect of AM is that monism was achieved even though the thesis of the anomalism or the irreducibility of the mental was taken to be true. In fact, since Davidson was “convinced” that the anomalism of the mental was true he could not, so to speak, allow the possibility of type–type reduction insofar as this requires laws connecting the mental and the physical. Such a reduction would have been in direct conflict with the thesis of anomalism. Given Davidson’s conviction that the anomalism of the mental is true, but so is physicalistic monism as well, one purpose of “Mental Events” was to “lay rest the view, common to many friends and foes of identity theories, that support for such theories can come only from the discovery of psychophysical laws.” This is an important claim insofar as it opens the possibility of O-physicalism without reduction. Those interested in defending physicalism should see this as a welcome result because reductive accounts of the mind have failed, and this raises the question of what reasons there might be to believe that O-physicalism with respect to mind is true. In other words, as long as successful mental to physical reductions are lacking, why believe that the mind is part of the physical reality in any substantial sense at all?

Davidson’s view became one example of a theory which is meant to be a version of physicalism, but which is also meant to be non-reductive. In the subsequent discussion, AM is often referred to as the locus classicus of non-reductive physicalism. Other well known forms of non-reductive physicalism are functionalism and emergentism, although the latter’s status as a version of physicalism has been seriously doubted and debated recently. What these

---

247 See Davidson, 1963.
positions share, according to the standard interpretation, is the view that mental properties will not reduce to basic physical properties. For this reason they are often described as forms of property dualism. Although these positions are seen as dualistic with respect to properties they, at least AM and functionalism, are meant to be versions of physicalistic monism since they hold that mental phenomena are token-identical with physical phenomena. With AM and functionalism the terms token-physicalism and non-reductive physicalism became part of the philosophical jargon, and they are usually referred to as the views which discredited type–type identity theories by showing their defects.

Let us briefly conclude what the importance of AM has been and what its place in the contemporary philosophy of mind is. According to Place, philosophers used to think, before identity theories were common, that materialism would not be the answer which solved the mind–body problem. Then, for somewhat unknown reasons, the opinion changed and for a short period of time, either type–type identity or eliminative materialism seemed to be a correct description of the mind–body relation. This period came to an end when AM and functionalism challenged type–type theories. Token-physicalism became the orthodox view. Consequent interest in emergentism and explorations of a more direct move away from physicalism are, for some at least, welcome side-effects of the success of non-reductivism. However, recently the situation has been changing again. The coherence of non-reductive physicalism has been under serious debate ever since the position became part of the philosophical discussion. A philosophically motivated identity theory has received new support from Polger. In the works of the Churchlands eliminative materialism has been vigorously defended again. John Bickle, a “new-wave reductionist”, defending a thoroughly physicalistic approach to the mind, has noted that: “Much current ‘non-reductive physicalism’ is not physicalism at all.” In the case of Davidson Bickle contends that some of his central remarks fit better with the idea of dualism than physicalism. Similar remarks about Davidson’s position have been made by Kim, who has become an important critic of non-reductive physicalism and has been identified as one of the modern supporters of reductivism. Kim argues that “robust physicalism” cannot tolerate a non-reductive position. His general message is: “Physicalism cannot be had on the cheap.” At the same time it should be noted that Kim’s latest book, Physicalism or Something Near Enough shows signs

249 The difficulties of the claim that mental properties are not physical properties and the coherence of the expression “property dualism” will be considered in section 2.3.
252 For Kim’s views about Davidson see Kim, 1989, 1998.
253 Kim, 1998, 120.
according to which also he is willing to admit that a thoroughly physicalistic view of the mind is not a plausible view. According to David Chalmers’ interpretation Kim actually turns out to be that of a “closet dualist”\(^{254}\). Kim’s exact position is not important here. What should be noted is that although the interest in reductive physicalism or eliminative materialism has seen a rebirth the situation seems to be changing constantly, and how well physicalism reflects the current consensus view is therefore somewhat unclear after all. The very basic aspects of different views about the mind are undergoing a lively discussion, and philosophers are trying to find out who defends what kind of position and why.

Despite the fact that AM challenged type–type theories, provided a novel argument for physicalism, and offered an interesting answer to the Kantian challenge, it is not a much supported view today. This is surprising because the three achievements were certainly not minor. The historical importance of AM should be obvious. Many would nevertheless claim that Davidson’s real philosophical legacy is not his contributions to the philosophy of mind but his achievements in the philosophy of language.\(^{255}\) I think this seriously downgrades the relevance of Davidson’s views. In the following I will turn to the details of AM before forming judgments about its contemporary relevance.

### 2.1.2 The physicalistic monism of AM

Recent general criticisms against non-reductive physicalism raise the question of whether the monism of AM should be thought of as physicalistic monism at all. The main criticism against non-reductive physicalism is, as Bickle’s view shows, that non-reductive physicalism is “not physicalism enough”. It is not robust physicalism because, although the non-reductive position accepts O-physicalism, it rejects M-physicalism. Some non-reductive physicalists are not moved by this charge because they are perhaps best thought of as antiphysicalists in a very strong sense.

There is a tendency to interpret also Davidson as a true antiphysicalist. His position has been given many labels. Bickle, for example, writes that there is an “antiphysicalist vein”

---

254 See the web blog of Chalmers: http://fragments.consc.net/djc/2005/09/jaegwon_kim_com.html

255 This impression I got from a personal discussion with Ernest Lepore and Barry Loewer, of whom the first one at least knows Davidson’s work very well. This obviously shows only the personal views of these philosophers, but what is telling is the fact that whereas there have been recently many publications about Davidson’s philosophy of language, there have been no interesting works on his philosophy of mind. Lepore is among those who have been writing about Davidson’s philosophy of language while ignoring his views in the philosophy of mind. See Lepore and Ludwig, 2005 and 2007.
in Davidson’s thought. Against this claim I want to emphasize that an important motive behind Davidson’s theory of mind was to secure the status of physicalism, and in this respect AM can be seen as part of philosophical accounts of the 1960s and 1970s which were used to defend physicalism. When Davidson is described as an antiphysicalist, a question that needs to be asked is whether he is trying to argue against physicalism. There are after all contemporary philosophers who try to defend a substantial form of dualism by arguing against physicalism, and these philosophers would certainly count as antiphysicalists on any interpretation. In my opinion the correct answer with respect to Davidson is that he argues against certain understandings of what physicalism requires or implies. I suggest that the argument structure should be seen in the broader context of Davidson’s philosophy as an attempt to defend certain very basic intuitions about ourselves, while at the same time criticizing some of the abstract demands that the contemporary naturalistic program has set for philosophy. The questions of what physicalism requires or what it implies are connected to the question of the reducibility of the mental. The reductive understanding of physicalism, M-physicalism in its various forms, is currently making a strong comeback. According to this understanding, physicalistic monism requires either the definitional or nomological reduction of higher level phenomena, or at least an explanation of them in physical terms. For those phenomena that have not yet been reduced, O-physicalism implies their reducibility or elimination. These requirements are expressed in different ways in the views of Fodor, Churchlands and other naturalists.

The view that questions of ontology must turn on the possibility of reduction is widely shared among naturalists. The possibility of reduction and its success dictate the right ontological status of the mind, i.e. whether it is reducible, eliminable or emergent. For a serious naturalist, emergence is not a real option because emergent phenomena would amount to nomological danglers which do not fit to the scientific picture of the world. These phenomena would amount to a mystery, which cannot be tolerated on this perspective. When natural science is given the role of ontology-maker the possibility of emergence has to be rejected, because the phenomena that a non-reductive physicalist or dualist would count as emergent would turn out to be candidates for elimination from the perspective of a reductive

---

258 For a general discussion about emergence and problems relating to it, see Beckermann et. al. 1992. What needs to be emphasized is that emergence is a live option for a classical naturalist, among whom I include contemporary thinkers like Rorty and Davidson. It is not an option for those contemporary naturalists who advocate a strict form of naturalism guided by the demands of science.
physicalist. Many see that the merit of AM is that it shows how a physicalistic monism is possible without reduction. In this sense, it should be counted as a very important contribution for physicalism. Björn Ramberg, a commentator whose interpretations Davidson appreciated, notes that one of Davidson’s great achievements is that he has shown how to dissociate reduction and ontology, thereby reducing the philosophical significance of both.259 Ramberg and others260 think that Davidson is, above else, offering a description of how it is possible to be a naturalist with “good conscience”, that is, without being a reductionist. Many naturalists worry that this is not possible. Fodor, for example, claims: “It’s hard to see… how one can be a Realist about intentionality without also being, to some extent, or other, a Reductionist.”261 This is the central claim of M-physicalism and, as I see it, it describes an absurd contention: something is not real if it cannot be reduced to the specific “natural” order that naturalism describes. Davidson does not subscribe to this form of physicalism or its ideology. The main reason for this is his conviction that the truth of physicalism can be established without reduction.

The disagreement between reductive and non-reductive physicalists reflects a more general disagreement about the nature of the philosophical method and about the relationship between philosophy and science. This is one of the reasons why I have explicated Davidson’s and von Wright’s conceptions of philosophy by comparing them to the ideology of contemporary naturalism. Rorty’s conclusion is that Davidson’s work is “a culmination of a line of thought in American philosophy which aims at being naturalistic without being reductionistic….”262 He sees Davidson as a pragmatist who has displayed no interest in a physicalistic ontology. Rorty thinks that this most clearly distinguishes Davidson from Quine, since the latter was explicitly anti-pragmatist in propounding a physicalistic ontology. That Rorty sees Quine’s commitment to a physicalistic ontology as the main difference between Davidson and Quine is interesting given that Quine came to accept AM. If both Quine and Davidson are anomalous monists, I cannot see how there could be a difference in their respective ontologies. I think Rorty’s statement that Davidson has not displayed interest in a physicalistic ontology is unclear, since it is obscure how this kind of interest is meant to be displayed. Rorty’s claim is suspicious also because the motivation behind AM was to give an argument for physicalism. Davidson has noted that: “Anomalous monism says that mental

261 Fodor, 1987, 97.
entities add nothing to the furniture of the world that is not treated in physics.”

He states also: “[...] my argument for anomalous monism... is designed to show that the ontology of any science that is not reducible to physics shares its ontology with physics.”

The latter point captures the essence of Davidson’s argument: all sciences and, indeed all “levels of reality”, share their ontologies with physics. Physics is thus ontologically basic. This, together with Davidson’s view that “perfect explanations” are those given in physics, surely raises the question of whether this does not show, pace Rorty, a bias towards a physicalistic ontology. According to Rorty, Davidson should be seen as a pragmatist. A pragmatist “should not say... that physics tells us all we need to know about the way the world is, but rather should follow Goodman and Putnam in saying that there is no 'Way the World Is’”. Davidson certainly denies that physics tells us all we need to know about the world but, as will become obvious, some of his views suggest that physics nevertheless can describe, and does describe, the “Way the World Is”.

Davidson has offered an argument which should dispel naturalistic worries, and in this sense he should be thought as contributing to secure the status of O-physicalism. Is this a relevant contribution? Fodor has claimed that there exists a “naturalist consensus” according to which: “something has to be said about the place of the semantic and the intentional in the natural order....” It is clear that Davidson does say something important about the place of the intentional in the natural order. In fact he says something which is philosophically very important. On the one hand, Davidson claims that any higher level science shares its ontology with physics. On the other hand, he claims that although mental phenomena are part of the natural order, they cannot be strictly reduced to that order as described by physics. Why shouldn’t this count as a positive contribution to the mind–body problem? It is a positive contribution, since it says something specific about the nature of the mental and the physical.

Davidson’s wants to show the truth of monism and dispel the common-sense (if such it is) intuition that mental phenomena will not reduce, in the ontological sense, to the natural order. But Davidson can relieve this “naturalist guilt” only if his philosophical argument for

---

264 Davidson, 1999a, 619. This is a curious view if the irreducibility to physics is principled.
265 Rorty, 1994, 126. As I see it, Rorty’s attempt to have Davidson as an ally in the pragmatist movement is the reason why he suggests that a clear distinction between a Quinean physicalistic and anti-pragmatist ontology and a Davidsonian ontology should be made.
266 Fodor, 1990, 12.
267 The only reason why it would not count as a positive contribution is if the abstract demands of contemporary naturalism are accepted and some descriptive systems are privileged over others due to metaphysical reasons.
268 The term is from Ramberg. It is the guilt that some philosophers “suffer as we continue to attribute intentional states before they are properly regimented through a psychology which lines them up with respectable predicates of natural science.” (Ramberg, 1999, 602)
monism succeeds. If there is no need for a philosophical argument for monism, I think one can accept, for example, Searle’s biological naturalism, which is described as a form of “scientifically sophisticated common sense”. According to this view mentality does not need naturalizing because, being a “biological phenomenon”, it is already part of nature. The mind–body problem has a simple solution. It is this: “Mental phenomena are caused by neurophysiological processes in the brain and are themselves features of the brain.” Astonishingly, contemporary philosophers have missed this point. It seems to me that if Searle’s view is enough for monism then philosophers should have turned into biological naturalists already after reading Darwin. Some of course did but many did not. An illustrative example is Huxley, a faithful follower of Darwin, who was nevertheless deeply troubled by the mystery of consciousness.

It is difficult to interpret, pace Bickle and others, Davidson as an antophysical because AM is meant to show that mental entities add nothing to the furniture of the world that is not treated in physics. This view clearly suggests that a substantial form of emergentism is not a consequence of AM. In fact, since physics is given an ontological privilege, there is a sense in which it can be said that micro-physical entities are more basic than others, at least in the sense that everything is composed of them, and their behavior governs the behavior of the entities they constitute. The actual physical story is obviously not simple, but this compositional picture of reality suffices as a very broad description of the privileged status of physics. On this reading Davidson’s physicalism, in its ontological respect, resembles Quine’s position.

One of Davidson’s important contributions to the philosophy of mind was the introduction of the concept of supervenience. Most commentators trace Davidson’s first use of this term to “Mental Events”. In fact Davidson entertained the idea already in his paper

---

270 Searle, 1992, 1.
271 Certain philosophers (e.g. Kim) claim that Searle’s oversimplified strategy cannot be thought of as a serious solution to difficult philosophical problems. I believe that Searle’s main point can nevertheless be appreciated. If contemporary philosophers believe in science, then something like biological naturalism ought to be a live option. Reductive and eliminative materialists criticize non-reductive physicalists on the grounds that this position is “unscientific”. Yet, from the perspective of biological naturalism, it can be claimed that both non-reductive and reductive physicalism, and all the problems which they encounter are a result of certain basic mistakes, above all the mistake of still thinking of the mental and physical in Cartesian, that is, in opposing or mutually exclusive terms.
272 Cf. Rorty (1998, 394): “[…] to make a big deal out of the psychological-physical distinction is a hangover of a early, pre-Darwinian epoch in the history of philosophy – the epoch of Kant.”
273 Since Davidson does not express his views about the nature of the actual theories of physics, I am not going to speculate about what his view about the nature of physics was.
“On Mental Concepts and Physical Concepts” from 1964 without using the term. The basic idea of supervenience, at least of its modern version, is the dependence of higher level properties on lower level properties, and of the determination of the former by the latter. Discussions of supervenience have been notable in the recent literature and the notion is used in a variety of ways. The subsequent discussion of this concept and the ways it has been used in the contemporary philosophy of mind gives, in my opinion, the impression that the introduction of this concept to the discussion was perhaps a mistake. Supervenience has turned out to be a mysterious relation which raises more questions than it solves. David Lewis notes: “We have supervenience when there could be no difference of one sort without differences of another sort.” But why such a correlation holds is usually left as a mystery in the discussion; supervenience has become a relation which simply holds, but for which there is no explanation. It has become a very technical notion and in the current discussion philosophical insights deriving from its use are few. The way in which supervenience has become an important notion in the contemporary discussion provides one example of how certain philosophical problems are treated in the spirit which is blind to any “Wittgensteinian” insights. The suggestion that philosophers should look to how words are used and learn from that has been completely ignored in these discussions. This makes the discussion and the “philosophical industry” that is continuing it quite frustrating.

The way in which Davidson has used the concept of supervenience suggests that he thinks that the basic physical facts determine, in some rather unexplained sense of the word, all higher level facts as well. Davidson writes:

> Although… psychological characteristics cannot be reduced to the others… they may be (and I think are) strongly dependent on them…there is a sense in which the physical characteristics of an event (or object or state) determine the psychological characteristics.…

Given Davidson’s claim that all sciences share their ontologies with physics, he thus endorses the metaphysical thesis that, ultimately, all facts supervene on the facts of the most basic of the sciences, physics. Supervenience implies monism, as Davidson concludes. It assumes mental–physical identity; so, as far as I see it, supervenience cannot be used to defend such identity and the argument for physicalism has to be established independently. Davidson’s use

---

274 Davidson, 1964.
275 For a discussion about the various forms of supervenience see Kim, 1993a.
276 For this complaint, see Schiffer, 1987 and Kim, 1987.
279 Davidson, 1993d.
of supervenience nevertheless shows how, in the context of AM, the relation between the mental and the physical should be understood; as a dependence relation of the former on the latter. This dependence or determination implies a mild form of physicalism since, according to Davidson, it is “obvious” that supervenience does not entail reducibility. This obviousness is nowadays strongly questioned, especially by Kim. His latest suggestion is that the best way to understand supervenience is to see it precisely as a reductive relation.\(^{280}\)

Although Davidson has made a contribution to physicalism, his attempt to show that O-physicalism can be achieved without M-physicalism is something that amounts to dualism for many. A contemporary naturalist argues that one cannot be a conceptual or methodological dualist and an O-physicalist at the same time. Fodor claims: “Psychologists have no right to assume that there are intentional states unless they can provide… naturalistic sufficient conditions for something to be in an intentional state.”\(^{281}\) Intentional realism requires showing how mind, as a material object, can have the properties it has. Davidson has not shown this, according to critics. In my view this criticism misses the point that Davidson’s purpose is to show the implausibility of any mind–body reduction. This reply does not move naturalists, who would complain that although AM, by establishing monism, perhaps very loosely integrates mentality into a physicalistic world view, it is by no means a satisfying theory of mind because it does not say anything specific about how the mental aspects of humans are related to their bodily nature. Perhaps the relation is accidental or a result of pre-established harmony. Although one could have thought that Davidson’s use of supervenience already shows a substantial commitment to physicalism and makes a suggestion about the nature of the mind–body relationship, many have argued that supervenience is a problematic concept through which to express this kind of commitment – precisely because it tries to combine determination or dependence with non-reducibility.\(^{282}\) Supervenience is thus a perfect, yet incoherent and therefore useless notion for a non-reductive physicalist. Many, Davidson among them, trace the notion of supervenience back to the writings of G. E. Moore. But as Bickle, who criticizes supervenience as a satisfying relation for a physicalist, notes: “Moore was unabashedly a dualist about descriptive and normative properties. Goodness supervenes on natural properties but is nonetheless – and, I might add, thereby – a nonnatural property.”\(^{283}\) Reductive naturalists claim that supervenience, far from providing support for robust physicalism, actually shows a commitment to a sort of dualism. The solution it

\(^{280}\) See Kim, 2005.
\(^{281}\) Fodor, 1994, 5.
\(^{282}\) Kim is the best known critic holding this view. See, Kim, 1989b.
\(^{283}\) Bickle, 1998, 9. How well this captures Moore’s actual view could be debated.
suggests for the mind–body problem is negative; supervenience merely states the way in which the mental and physical natures of humans are not related. As a curiosity I want to note that Moore never used the term supervenience, although he of course uses the notion (which Davidson, among others, attributes to him). It was in fact Richard Hare who in his *The Language of Morals* from 1952 introduced the term to general philosophical public.\textsuperscript{284} Hare himself notes, however, that he did not originate the term but was only following a use which was already familiar. It is possible that he is referring to the article “On Grading” by J.O. Urmson, which was published in 1950.\textsuperscript{285} Urmson writes: “Having rejected naturalism, but recognizing the close connexion between X and A B C, we shall say... that X-ness... is a non-natural, intuitable, toti-resultant character supervening on situations in which A B C are present, necessarily, but synthetically, connected with A B C.”\textsuperscript{286} Whatever the exact history of supervenience, it is interesting to notice that Hare, who is often seen as the figure behind supervenience, criticizes Davidson on the grounds that he has turned a perfectly clear and non-mysterious relationship into a mysterious relation.\textsuperscript{287}

Some M-physicalists argue that since everything is physical, they are entitled to say that all apparently “non-physical” facts are completely determined by physical facts and, this being the case, they are nothing “over and above” physical facts. This is firmly denied by Davidson, since he claims that even a mature physical science could not exhaustively explain the essence of the mental. This is one important sense in which mental phenomena, although identical with physical phenomena, are something over and above physical phenomena: they cannot be given purely physical explanations. This would be denied by most M-physicalists. Although many of them come out as non-reductivists because they reject the possibility of type–type reduction, or even actual token-token reduction, they nevertheless insist that some kind of explanation of mentality in non-mental terms is required for the vindication of mentality. In the words of Fred Dretske, the question is: “Can you bake a mental cake using only physical yeast and flour?”\textsuperscript{288} If you can’t in principle, then their conclusion is the elimination of the mental.

\textsuperscript{284} Davidson (1993) acknowledges that he probably picked the concept form Hare, although he earlier (see for example Davidson, 1964) traced the source back to Moore.
\textsuperscript{285} Urmson, who was working in Oxford, would be one natural candidate for the source because Hare recalls that the term was used in Oxford during the 1940s although he does not remember in what context and by whom. On the other hand, as van Brakel (1990) notes, the concept was used already in the 1920s by many as a somewhat synonym for “emergent” and this being the case it is possible that both Davidson and Hare may have picked the concept up from these contexts. If Davidson picked this concept up from “emergentists”, it would be an interesting question how he originally saw the nature of the relation it expresses.
\textsuperscript{286} Urmson, 1950, 155.
\textsuperscript{287} Hare, 1984.
\textsuperscript{288} Dretske, 1981, xi.
It is understandable that from the perspective of robust M-physicalism, Davidson can perhaps be described as being “antiphysicalist” and a property dualist. This accusation of dualism is based on the conviction that physicalism and non-reductivism cannot co-exist. I think that one way to explore the coherence of this position is through a consideration of whether Davidson’s argument for monism succeeds. There are also other reasons to accept the compatibility of non-reductivism and physicalism, but my interest is to find out whether Davidson’s argument for such a position is satisfying because the purpose is to evaluate the coherence of non-reductive physicalism as understood by Davidson and von Wright. In Davidson’s case, the dilemma is that if his argument for monism succeeds, it is because the irreducibility of the mental has been established. In the context of a monistic world view, as in the picture described by naturalism, predication and explanation are important notions. The extent to which prediction and explanation succeed show how well the nature of reality has been grasped. In this context, Davidson’s claim that human behavior cannot be explained or predicted in the same way as everything else amounts to substantial dualism, to a form of mysticism. From the perspective of M-physicalism, this position is dualist because it locates the mental aspect of humans on a wholly different level than everything else. This kind of ineliminable irreducibility would mean that mental phenomena would end up being nomological danglers, just as the type–type theorists of the 1950s had feared.

In the next section, the physicalistic content of AM is discussed. In this section I have presented textual evidence showing that AM is, or is at least meant to be, a version of O-physicalism. How Davidson himself saw his position is relevant when considering what the nature of the position in fact is, and how it should be understood. I think it is important to consider the textual evidence, because AM has been interpreted also as a form of neutral monism. Davidson’s position can be described as a form of physicalism or a form of dualism, and it is important to see the reasons for both interpretations. My view is that it is hard to challenge the assertion that Davidson argues for a physicalistic monism given his claims that: “[…] there is only one set of entities in the world”289 and all of its members are physical. What something’s being physical means is a question for the next section, but one should note that Davidson is searching for quite a strong form of physicalism, one which lies somewhere between mere O-physicalism and strong M-physicalism. The argument for AM is meant to show how this kind of metaphysical view is possible. In “Mental Events” Davidson stated his position by saying:

289 Davidson, 1963, 46.
Anomalous monism resembles materialism in its claim that all events are physical...Anomalous monism shows an ontological bias only in that it allows that not all events are mental, while insisting that all events are physical.\textsuperscript{290}

This is a delicate formulation. AM \textit{resembles} materialism because it insists that all events are physical and allows that there are non-mental events. However, the monistic point is made more strongly in “The Material Mind”, where Davidson writes: “psychological events simply \textit{are} (in the sense of \textit{are identical with}) physical events. \textit{If this is materialism, we are committed to it…}”\textsuperscript{291} In “Psychology as Philosophy” it is claimed that “[...] psychological events are describable… in physical terms, that is, they are physical events… this position deserves to be called… monism, because it holds that psychological events are physical events…”\textsuperscript{292} Finally, a clear statement which leaves no room for non-physicalistic interpretations: “Monistic my view is, since it holds that mental events are physical events.”\textsuperscript{293} These statements point toward physicalism in a straightforward sense; they express an “ontological bias”. This will not surprise those who are familiar with Davidson’s philosophy. Yet, for two reasons it is important to emphasize that Davidson is trying to \textit{defend} ontological physicalism and explicate the nature of it.

First, some commentators have suggested that Davidson is arguing for a kind of neutral monism which claims that events in themselves are not physical or mental. Thus one commentator, Louisa Röska-Hardy writes: “What exists is non-abstract particulars…. They are neither mental not physical in themselves....”\textsuperscript{294} In a similar spirit, Marc Joseph claims: ”Davidson eschews the mental and physical as ontological categories”\textsuperscript{295} and “the natural cuts across the mental and the physical, which are conceptual (and not ontological) categories.”\textsuperscript{296} This view agrees with the suggestion that Davidson has tried to show how one can be a naturalist without being a reductionist.\textsuperscript{297} The view is a correct description of the Davidsonian position insofar as it identifies “the natural” as an all-encompassing category which cuts across the mental and the physical; yet, or so I believe, the claim misfires by attributing to Davidson the view that the physical is a conceptual category. Simon Evnine concludes with respect to Davidson’s view that:

\begin{itemize}
  \item Davidson, 1970, 214.
  \item Davidson, 1973, 248, second emphasis mine.
  \item Davidson, 1974b, 231.
  \item Davidson, 1985, 245.
  \item Röska-Hardy, 1994, 291.
  \item Joseph, 2003, 152.
  \item Joseph, 2003, 163.
  \item Ramberg and Rorty, among others, make this claim. The claim would fit with my view that Davidson’s position comes close to classical naturalism.
\end{itemize}

104
[...]

But how could these interpretations be taken seriously? To say that “physical” is not an ontological category or that Davidson has chosen to call his position “monism” rather than “materialism” certainly seems to conflict directly with the textual evidence that I have presented. To say, as Evnine does, that it is not “important” to claim that all events are physical seems to be a serious misinterpretation concerning Davidson’s intentions. Davidson’s events are not “ontologically neutral”, as Evnine seems to think.

The second reason why Davidson’s purpose should be emphasized is this. One should be interested in Davidson’s argument for physicalism as long as the truth of ontological physicalism is not taken for granted. Many contemporary philosophers accept physicalism because they simply agree that the ultimate ontology should be continuous with the current scientific ontology. Davidson’s philosophical, a priori, argument for physicalism is interesting, because it gives an additional reason to think that physicalism is, or could be, true. In my opinion it is curious that Davidson’s statements emphasizing the physicalistic aspect of AM are, more or less, from the same time period. At that time, the need to stress the physicalistic content of AM seems to be at the forefront. The reason for this is perhaps the fact that, as I have shown, AM was meant to be an argument for physicalism, and it was meant to back up other identity theories of that time. Whereas it is important to consider the correctness of AM, it is also important to place AM to the context of its time. In the 1970s there was a need for a physicalistic view about the mind, a need which arose from the fact that the truth of physicalism was still under debate although the empirical, if not philosophical, evidence had started to turn in its favor.

There is thus clear textual evidence that allows the conclusion that Davidson’s position is a form of monism, because it holds that all events, or more broadly “entities”, are physical. It is usually emphasized that Davidson’s physicalism concerns, above else, events. I extend the discussion more broadly to entities, because Davidson himself has, from time to time, used this expression. He writes, for example: “[the premises of anomalous monism] imply that mental entities do not add to the physical furniture of the world.”

Although Davidson officially claims that he is defending only the view that mental events are physical events, I

---

298 Evnine, 1991, 64, my emphasis. A similar view is suggested by Leonardi, 1999.

299 Davidson, 1994a, 231.
think it is clear that the spirit of his position aims for an all-encompassing metaphysical position. O-physicalism claims that there is only one set of entities in the world, and this claim is accepted also by Davidson. Physicalism is the important philosophical and metaphysical thesis to which we are committed as a conclusion of Davidson’s argument. Depending on one’s other views about what there is, the furniture of the world may or may not include entities like states, processes, and properties. In addition to mental events, Davidson accepts at least the existence of mental objects – the nature of which remains unclear in his writings. It seems to me that his position, insofar as it accepts, for example, processes, must hold that also these entities are physical. In the following I shall use the expression mental phenomena (which includes everything that is intuitively called mental). What I want to determine is whether Davidson and von Wright give reasons to believe that these phenomena should be thought as physical.

Given that Davidson presents a non-empirical philosophical argument in favor of physicalism, it is not surprising that he describes the result as being a philosophical thesis. This certainly differs from the view of, for example, Place who thinks that the type–type identity theory is a “scientific hypothesis.” The view that physicalism is a scientific hypothesis is generally shared among contemporary naturalists. The wide acceptance of this postulate seems to have granted physicalism almost something like an a priori status in the current discussion. It has become like a necessary feature of reality, even though its origin is empirical.

2.1.3 Davidson’s physicalism

So far I have claimed that AM is (meant to be) a form of physicalistic monism. It belongs to the category of physicalistic accounts of the mind, and solves the traditional mind–body problem by arguing for monism and suggesting that mental phenomena (Davidson’s “mental entities”) are token-identical to physical phenomena. Davidson’s answer to the traditional mind–body problem is thus clear: Mental and physical phenomena are identical; ontologically speaking, whenever one speaks of something mental there exists only something which is physical. But what does it also mean to claim that reality contains only physical entities? The answer is not easy to give, because the exact nature of Davidson’s physicalism is unclear. What does Davidson mean when he argues that everything is physical? I believe there are two

---

300 See, Davidson, 1974b, 248.
301 Place, 1960, 1989.
ways to understand the claim that there exists only one kind of phenomenon. These two ways are intertwined in a confusing manner. The two alternatives are an ontological reading of physicalism and a conceptual reading of physicalism respectively.

Physicalistic monism claims that every event and every substance is physical. This is not an illuminating clarification unless one is told what something’s being physical means. Noam Chomsky has claimed that as soon as we come to understand something we call it physical. This is a perspicuous formulation which captures the scientific mood of our time; yet it obviously leaves out a detailed description of what being physical means: it does not give a criterion of the physical. The criterion of the physical has been difficult to define, and currently there is no definition which is generally accepted among philosophers. It is also clear that scientists doing research in physics or in other physical sciences have different views about the nature of the physical from philosophers who are searching for a general criterion. However, most contemporary physicalists (that is, philosophers) refer to physics when judging what something’s being physical should be taken to mean, and therefore see physics as the final arbiter for declaring what should or can count as a physical entity. This conclusion rejects the need for further philosophical attempts to define the meaning of “physical”.

The view that science, or more specifically physics, tells us what there really is, is a result of the naturalistic attitude which can be summarized in Quine’s words: “The world is as natural science says it is….” The consequence of this, at least according to some interpreters, is that: “If natural science cannot tell us more about how [something] works, then it does not exist.” It is not surprising that many, von Wright among them, have seen modern philosophy as being partly in the grip of scientism. According to many naturalists, physics is the purest of the natural sciences because it attempts to describe the most basic level of reality. What really exists is a question for physics. In Quine’s case, for example, his “[…] ontology continues to consist of quarks… and the like.” Kim has recently noted that Quine’s view – according to which physics is, by definition, the science which aims at full coverage – has interesting implications for what it is to count as physical. By definition, nothing can lie outside the reach of physics. The sphere of “physical” could be extended if

---

302 As mentioned by Searle, 1994.
303 Unless one accepts that being physical = being understandable by humans.
305 Rorty, 1994, 124.
306 Ibid.
307 In an interview retrieved which can be found at http://www.ephilosopher.com/page.php?15
deemed necessary by physics, but our ontology – our catalog of what there really is – would still be continuous with the discoveries of the most basic part of physics.

If physics ultimately provides the criterion of the physical, it seems to follow that if mental phenomena are physical phenomena, then they should have a property which all physical phenomena have – and this most basic property would be whatever physics declares it to be. Given the current state of physics, it is unclear whether there in fact exists such a property. Let us suppose for the sake of argument that there is such a property – a property which entities must have if they are to count as physical – and physicists can tell us what this property is. Obviously, this property must be something which can change as physics develops. This means that in the future it may turn out to be the case that not all physical phenomena have this currently attributed property, or that a more fundamental property is discovered – a property which again all physical phenomena are required to have. Should we thus settle for the view that the criterion of the physical is what the physics of any given time declares it to be? Something like this seems to be Quine’s view, since he claims that ontological commitments should be continuous with the discoveries of physics, whatever they are. Whatever conclusions physics draws about the nature of reality, a situation in which our ontology would turn out to be non-physicalistic is not an option. Forces which would now count as obviously non-physical may be explained in terms of some future physics.

If we ignore the problem that the criterion of the physical is apt to change and settle for the view that our ontology should be continuous with the ontology of physics, then what can be said, given our current knowledge, about the nature of physical entities? It is often said that material (or physical) entities can be located in space and time. Perhaps a space-time location property could give one acceptable criterion that an entity must satisfy if it is to count as physical. This criterion goes well with the idea that physical entities should be observable or measurable and, in this sense, objective. They are located in a “common space” which, in principle, is accessible to all.\footnote{308 Macroscopic material objects are extended in space and they occupy the space in which they are. They can be observed objectively. It is not clear whether this is the case with abstract physical phenomena or with phenomena at the sub-atomic level.} Macroscopic material objects are extended in space and they occupy the space in which they are. They can be observed objectively. It is not clear whether this is the case with abstract physical phenomena or with phenomena at the sub-atomic level.

Whether or not spatio-temporal location is an acceptable criterion of a physical entity, modern science nevertheless puts forward a compositional picture of reality. Macroscopic objects are “made up of” smaller particles, which in turn are composed of even smaller particles until the base level of elementary particles like quarks and leptons is reached.\footnote{Intuitively mental phenomena are not like this. Their naturalization would therefore require that they be reduced to something which is observable by anyone.}
Perhaps this could be used as a criterion of the physical: everything physical that is composed of anything is composed of elementary particles. This seems to be true – at least with respect to those entities that can be thought of as being composed of something. Physicalistic monism, which endorses a compositional view of reality, ontologically privileges to the entities described by physics. Microphysical entities are fundamental entities, because everything else is composed of them – and the nature of higher level phenomena and their behavior, so one suggestion goes, could be explained in terms of the most basic entities by referring to laws which govern the behavior of these entities. How well this simplistic and rather mechanistic view reflects the current thinking in physics is unclear, but philosophers often use this kind of picture quite unhesitatingly. When we apply these considerations to the mind–body problem, it seems that in order for mental phenomena to count as physical, they must be composed of physical particles. This is what most modern philosophers of mind conclude. As a physical criterion of the mental, this is perhaps acceptable.

Physics is thus the final judge regarding the criterion of the physical. This view is not as interesting for the scientist as it is for the philosopher. Quine’s interpretation or definition of physics is that physics is “factual”, whereas inside other discourses there may be not a fact of the matter regarding what there is or what there is does. For Quine, the notion of such a ‘fact of the matter’ is an ontological one, “a question of reality”, as he says. The question of what there really is will be decided, as we have seen, via a scientific theory of the world. Although Quine has been somewhat ambivalent in his views about ontology, his position can be summarized by stating that the ultimate facts of the matter are decided in terms of the most basic elementary particles, or microphysical states, to use Quine’s favorite expression. Matters of fact will depend on the distribution of microphysical states, and these states will be identified according to the physics of the moment. To say that the states identified are microphysical states is merely a consequence of the fact that the states in question are identified by physics – which, by definition, gives the criterion of the physical. According to the Quinean view the “amount of factuality” – that is, the extent to which statements refer to extra-linguistic, non-human, physical reality – is greater the “closer” these statements are to physics. Physics is more factual than chemistry, which is more factual than biology, which is more factual than psychology and so on. The non-factuality of psychology forces Quine to draw eliminativist conclusions about the status of mental phenomena. This is one clear

---

309 Quine, 1981, 23.
example of the way in which the ontology of physics is privileged: higher level ontologies may face elimination if they do not smoothly reduce to the ontology of physics.

Given that a physicalist (philosopher) is committed to accepting the changing ontology of physics it is useful, instead of trying to give a criterion of the physical, to ask what it is that a physicalist claims about the world. The general claim of O-physicalism is clear: there is only one set of entities in the world. But a more detailed answer is bound to have more specific ontological consequences given that it is physics – the science – whose path a physicalist has to follow. According to Quine and Davidson, physics aims at “the full coverage” of reality. Quine claims that this has the following consequence: “If the physicist suspected there was any event that did not consist in a redistribution of the elementary states allowed for by his physical theory, he would seek a way of supplementing his theory.” This is a philosopher’s view about the situation; it is not obviously the view that a physicist would hold. If this advice of a philosopher is to be followed, the question is how the theory should be supplemented? No matter what the answer is, the modifications would still concern, by definition, microphysical states. But to say that physics has the aim of completeness and that whatever physics postulates is physical by definition is not a very interesting claim. It may be certainly something to which a physicalist is committed, but it is hardly a reason to believe that physicalistic monism is true. The problematic philosophical question has always been whether mental phenomena are physical phenomena. If they are and if the completeness of physics (all events are determined, or have their chances determined, entirely by prior physical events according to physical laws) is true, then there are no obstacles, in principle, for giving physical explanations of mental phenomena. If mental phenomena are physical phenomena, then a major – perhaps the most serious – obstacle for a physicalistic world view is removed. Then one is can agree with Quine that “nothing happens in the world, not the flutter of an eyelid, not the flicker of a thought, without some redistribution of microphysical states.” But this kind of supervenience assumes physicalistic monism, as was already noted by Davidson. It seems to me that talk about the supervenience of the mental on the physical – or of the former being “realized” or “constituted” or “determined” by the latter – all this talk is appropriate only after the general metaphysical assumption of physicalism is in place. But one first needs a reason to believe that physicalism is true. Once there exists such a reason, the details of a physicalistic world view can be discussed more seriously.

---

310 Quine, 1981, 98.
311 This formulation is from Papineau, 2001.
312 Quine, 1981, 98.
Davidson expresses the monism of AM by claiming that everything that exists can be given a physical description – and ultimately a description in the vocabulary of physics. The fact that an event can be given a physical description is what “makes” it a physical event. But I think we should ask why we should believe that everything can be given a physical description. A reason to believe that physicalistic monism is true would be a reason to think that, in principle, everything can be described in the language of physics. If one had a criterion of the physical which was independent of physics and if one had reasons to believe that mental phenomena satisfied this criterion, then one would have a reason to think that mental phenomena could be given a description in the language of physics if it was further assumed that physics can describe everything physical. It seems to me that the claim that physicalistic monism is true (and therefore everything can be given a physical description) and the claim that everything can be given a physical description (and therefore physicalistic monism is true) are just two different ways of expressing the position of monism. But, as far as I can tell, one would need an independent reason to believe either one of these claims. Describability in the language of physics can be taken to be a criterion of the physical, but whether or not everything can be described in this language is unclear from such a criterion. If something remains indescribable in the language of physics, it cannot be counted as a physical entity.

According to this view, it seems to be the case that if mental phenomena (the existence of which is a given for us) turn out to be indescribable in the language of physics, then they simply are not physical phenomena. This approach, according to which everything can be given a physical description, does not give an argument for physicalism; it merely states quite broadly what physicalism means. This is the approach of Quine, who accepts physicalism as the most plausible theory about the nature of reality. His answer to the question of why everything must be describable in the language of physics would thus be that this is the situation according to our best scientific knowledge; nothing can remain outside the reach of physics, which aims at full coverage of reality. Given Quine’s commitment to the ontology of science, it is unclear whether the kind of physicalism that he defends could be refuted in any way.

As I have argued, Davidson’s purpose is to give a philosophical defense of physicalism – and in this respect his approach is more philosophical than Quine’s, whose “argument” is his naturalism. Quine has no need for the kind of argument that Davidson is intending to provide, and he did not attempt to provide such an argument in his philosophy of mind. Davidson, on the other hand, says things like: “Each mental event has… (must have, if I
am right) a physical description…”\textsuperscript{313} or “… every… action… is also, if I am right, a physical event…..”\textsuperscript{314} This gives the impression that, according to Davidson, the truth of physicalism is something which has to be argued for, and that he has provided an argument. I think we should agree that if no philosophical argument is needed for physicalism, then mental phenomena should be counted among physical events. One can just accept biological naturalism. But Davidson’s argument is not based on the claim that the truth of physicalism seems to be a scientific truth, “a plausible view”, and should therefore be accepted. If one accepts Davidson’s argument, it follows that everything can be described in the language of physics. But it is not clear what this implies about the nature of the entities so described. If an entity can be described in the language of physics, does it then have a spatio-temporal location – and if so, should this then be the actual criterion of the physical? If an entity can described in the language of physics, then this – so we could argue – says something substantial about the entity. But according to Davidson, this – whatever physics says about physical entities – is not the final criterion of the physical; the criterion is that the entity can be described in the language of physics. In “Mental Events” Davidson asks: “What does it mean to say that an event is mental or physical?”\textsuperscript{315} The answer is: “[…] an event is physical if it is describable in a purely physical vocabulary, mental if describable in mental terms.”\textsuperscript{316} So, if an event or object has a physical description, it simply is a physical event or object. This is the criterion.

Some have argued that if the mere possibility of being amenable to a physical description makes an entity physical, then the resulting form of physicalism is so weak that it should be not called physicalism at all. Skillen claims that if events are physical “only as described” then Davidson’s notion of an event is “hopelessly minimal”\textsuperscript{317}. Skillen does not claim that Davidson would clearly subscribe to this minimalistic notion of an event. He would thus agree with those commentators who argue that, according to Davidson, there is a “substantial difference” between mental and physical events due to the fact that events are mental only as described, but physical intrinsically. Emiliani, for example, claims that: “The most striking consequence of Davidson’s view that ‘events are mental only as described… is that the very existence and nature of mental events is made language-dependent and

\textsuperscript{313} Davidson, 1990b, 92, my emphasis.
\textsuperscript{314} Davidson, 1999t, 335, my emphasis.
\textsuperscript{315} Davidson, 1970, 210.
\textsuperscript{316} Ibid.
\textsuperscript{317} See, Skillen, 1984, 523.
contrasted with the language-independent existence of physical events.” But what does the ‘language-independent existence of physical events’ mean? What, on Davidson’s view, is the criterion of the physical which does not depend on language?

Also Kim notes that Davidson’s view is an example of very weak form of token-physicalism, whereas other critics, like Latham and Antony argue more strongly that Davidson’s physicalism does not exclude the possibility of substance dualism – and thus cannot be considered physicalism at all. I think their criticism is sound, since it seems possible that an entity could get a physical description, thus becoming a physical entity, even though it could include (as a part of it) the thinking of an immaterial mind, for example. This is a fanciful example, but I think it makes the point. A substantial criterion of the physical is lacking if the property of being describable in a physical language suffices for something to count as physical. But perhaps one can settle for the view that whatever physics tells us about basic entities constitutes “what they really are”. A realist understanding of physics suggests such a view. If something can be described in the language of physics, then this something is really made of elementary particles, has a spatio-temporal location, and can be measured – whereas a statement that something is beautiful or good does not express any fact of the matter, although it is based on the interpretation of objective facts that depend on the nature of agent-independent reality.

Although Davidson is unsure whether a satisfying criterion of the physical can ultimately be given, there is one important passage where he says something essential about the nature of the physical. For something to count as physical, it must be describable in a “physical vocabulary”. But how is this physical vocabulary to be characterized? Davidson does it in two ways. On the one hand, he claims that physical entities are those entities which are describable in the vocabularies of the natural sciences. Sometimes Davidson refers to the language of physics, and sometimes he notes that “a particular physical event, state, or disposition is one that can be picked out... using a vocabulary from some physical science.” On the other hand, Davidson often uses the term “physical” in a wide sense, according to which all statements without intentional terms are physical. His use of the term “physical” involves an amount of indeterminacy. In one of his earliest articles Davidson writes that “physical concepts... are tied to the common sense notion of a physical object.

---

318 Emiliani, UP.
319 Kim, 2005.
322 Davidson, 1990b, 92.
which has a location in space and time, which undergoes changes which are physical events, whose changes are governed by laws, that is, physical laws.\textsuperscript{323} The point is that in order to discuss physical phenomena meaningfully, there must be some agreed principles which govern our physical vocabulary. We obviously mean and must mean something when we describe an entity as “physical”. In this important passage, Davidson suggests that physical concepts are tied to the “common sense” notion of a physical object. There is no mention of physics or any of the other natural sciences. But if it is the common-sense notion of a law-governed object in space and time to which our understanding of the physical is tied, what should we then say about the property that all entities share (supposing that there is only one set of entities in the world)? I would suggest that the most important characteristic of physical entities, according to Davidson, is that their behavior is law-governed. If the relevant laws were known, everything physical could be explained and predicted accurately. This, then, is the essence of the physical: it is strictly law-governed. Every physical entity and phenomenon has this property. However, this is not a criterion straightforwardly suggested by physics; it is not an independent criterion.

Since I am interested in Davidson’s reasons for adopting physicalistic monism, I shall not consider the many arguments which have been offered by other philosophers in favor of the view that everything is physical. As Papineau, Kim, Burge and others have argued, it is plausible to think that scientific achievements during the 20\textsuperscript{th} century created a favorable intellectual climate for physicalism. Therefore the amount of \textit{philosophical argumentation} for physicalism has been decreasing as science has revealed ever more details about the nature of reality. What is interesting about Davidson’s argument is that it is a philosophical argument; it is not directly based on empirical evidence and it is not based simply on the naturalistic view that our fundamental ontology should be dictated by science. Davidson’s argument is simple, and in my opinion it is clear that it fails. How does Davidson reach the conclusion that everything is physical? Monism is reached from three premises that we have already seen:

1) The Principle of Causal Interaction: At least some mental events interact causally with physical events.
2) The Principle of the Nomological Character of Causality: Events related as cause and effect fall under strict deterministic laws.

\textsuperscript{323} Davidson, 1964, 45.
3) The Anomalism of the Mental: There are no strict deterministic laws on the basis of which mental events can be predicted and explained.

Here I shall not discuss principles one and three in detail. Principle three will be discussed and defended in the following sections. Principle one will here be accepted for the sake of argument. In chapter four its correctness will be questioned.

Davidson’s argument for physicalism is based simply on the view that causal interactions between events must fall under strict deterministic laws. If principles one and three are true, then principle two is crucial for monism. But how do these three principles imply monism? The “demonstration of identity”, that is, the demonstration that each mental event is a physical event, is easy. According to the principle of causal interaction, at least some mental events interact with physical events. Whenever there is causal interaction, there must (according to Davidson) be a strict law covering the case, as the principle of the nomological character of causality claims. Since, according to the anomalism of the mental, there cannot be strict laws which involve mental terms, the law covering such a case must be a physical law. In order for an event to fall under physical law it must have a physical description, which, on Davidson’s criterion, means that it is a physical event. The demonstration thus shows that if a mental event causally interacts with a physical event, the former must have a physical description and must therefore be a physical event.324

Because there must be a strict law covering causal interaction, and given that there cannot be strict laws involving the mental, the lawlessness of the mental is used (of course together with the principle of the nomological character of causality) to establish the identity of the mental and the physical. What makes AM’s case special among identity theories is that whereas previous identity theories were based on the view that empirical investigation and evidence can provide support for the formulation of laws between the mental and physical (and are therefore needed for the establishment of identities), Davidson claims that a token-identity theory does not require such evidence – and in fact that the truth of token-identity theories actually depends on there not being such law-like relations. This raises very interesting questions about the relevance of empirical evidence for Davidson’s monism. Empirical evidence is not needed for Davidson’s argument – but if there were evidence

---

324 That a view about the nature of causation is the key argument in Davidson’s monism is studied in depth by Yalowitz, 1998.
according to which the mental realm is not law-governed, would this then strengthen Davidson’s argument for monism?\textsuperscript{325} This is a question worth considering.

The problematic principle of the nomological character of causality is crucial for Davidson’s argument. What are the problems involved with this principle? First of all, it is not an empirically verified principle. This raises the question of what the reasons to believe its truth are. Second, many think that current physics does not guarantee that causation is deterministic. Davidson may be thus be mistaken in his claim that strict laws are deterministic. Of these problems, the first is more serious. If the principle of the nomological character of causality cannot be shown to be true, the philosophical argument for monism fails and AM, as a physicalistic position, is doomed. What reasons are there to believe that singular causal statements are backed up by strict laws? Davidson notes that:

\[
\text{[…] if there is a reason for holding the cause-law thesis [the principle of the nomological character of causality], the argument must in some sense be a priori, for the thesis clearly is not a pronouncement of ordinary logic, nor can it be established empirically.}\textsuperscript{326}
\]

Davidson’s reference to an \textit{a priori} argument is interesting given what has been said about his conception of philosophy. The idea that there is a law, deterministic or not, which backs up singular causal claims must be given an \textit{a priori} argument. The task of a philosopher is to provide such an argument. But what is it? Many commentators have concluded that there is no reason to believe that the principle of the nomological character of causality is true and, if there are reasons to believe in its truth, Davidson has not offered them. Elisabeth Anscombe notes:

Davidson, will say, without offering any reason at all for saying it, that a singular causal statement implies \textit{that there is} such a true universal proposition [Always when this, then that] – though perhaps we can never have knowledge of it. Such a thesis needs some reason for believing it! ‘Regularities in nature’: that is not a reason.\textsuperscript{327}

A more recent observation comes from Robert Van Gulick:

That an event C of a given type should produce an effect E on one occasion and that another event C* physically indistinguishable from C in all respects should none the less fail to produce any event similar

\textsuperscript{325} Could there be such empirical evidence? Evidence according to which the brain would be some kind of chaotic system would not be helpful because, in order for this evidence to be relevant for premise three, the identity between mind and brain would have to be assumed beforehand. It is hard to think of any kind of scientific empirical evidence which does not assume mind-brain identity but shows that the mind is not law-governed.

\textsuperscript{326} Davidson, 1995c, 202.

\textsuperscript{327} Anscombe, 1975, 81.
According to van Gulick empirical evidence goes against Davidson’s view. Brian McLaughlin is explicit in his criticism: “Davidson has never offered any reason to believe the thesis. Nor... has he ever argued for the principle of the nomological character of causality.”

I believe these criticisms were in order at the time when they were presented. As a reply to this kind of criticism in “Laws and Cause”, published 25 years after “Mental Events”, Davidson explains his reasons for believing the cause-law thesis. I have to say that his reasoning in this paper is unfortunately quite obscure and confusing. In the subsequent discussion, commentators have not stated whether the argument convinces them. It seems to me that the argument has not convinced the critics. Kim has recently noted, repeating McLaughlin’s earlier complaint, that “Davidson never stated a clear reason, much less a detailed argument, for his requirement of strict laws for causal relations.”

Walter Freeman claims that developments in physics have shown that brains are not material systems for which the laws of physics support accurate prediction. According to him, this empirical fact undermines AM. Freeman writes that brains are “capable of self-organizing chaotic dynamics that lead to unpredictable and complex new behavior.” I cannot evaluate the correctness of this claim, but if there is evidence in its favor then this would seem to undermine Davidson’s reasons for both monism and for the anomalism of the mental. If the brain is a physical system, the behavior of which cannot be strictly predicted, then the argument that all causal relations between physical events are governed by strict laws is false. If the brain is capable of producing unpredictable and complex new behavior, then the argument according to which mental events cannot be predicted is false. Empirical evidence thus threatens Davidson’s argument structure. What makes things worse is the critics’ suspicion that a conceptual argument for the principle of the nomological character of causality cannot be given. This would, of course, seriously undermine the monism of AM.

---

328 Van Gulick, 1993, 239.
330 Davidson himself has noted that in the last part of “Laws and Cause” he was in “way over his head”. See his comment in an interview at: www.stanford.edu/group/dualist/vol7/pdfs/davidson.pdf
331 Kim, 2007, 229.
332 Freeman, 1999, 2006. Freeman does not claim that brains are not material systems. He claims that they are systems for which the laws of physics do not support accurate predictions.
333 Freeman, 1999, 145.
Davidson nevertheless claims, along Kantian lines, that it is a conceptual truth that if a singular causal statement is true, then there must be a covering law which is strict. Here we should again note Davidson’s views about the nature of philosophy. Philosophical truths are results of conceptual investigations; perhaps the best philosophical conclusions amount to conceptual truths. According to Davidson, accepting the conceptual truth for which he argues does not amount to accepting the regularity of nature. So, I think that against Anscombe Davidson can reply that “regularities in nature” are indeed not the reason for the cause-law thesis. But what does it mean to say that the acceptance of a conceptual truth does not amount to acceptance of the regularity of nature? Davidson gives the impression that physics discovers objective facts about reality; the regularities of physics are objective regularities of nature. This is a realistic understanding of physics, and there are no good reasons to doubt that Davidson would not subscribe to it. If the connection between laws and cause is conceptual, how is this claim used to establish physicalistic monism? The monistic view is about the nature of reality; it is a claim about the ontological constitution of the world and should be separated from epistemological considerations. But how can a conceptual truth establish something about the nature of the extra-linguistic reality?

The conceptual nature of the cause-law thesis has been challenged, for example, by Armstrong and Heathcote. They suggest:

Consider a world that is a world of causes and effects—where, perhaps, every event has a cause and in turn is a cause—but where causes of the same sort do not give rise to the same sort of effects. It sounds quite wrong. But can it be shown purely by analysis of the concepts involved that such a world is impossible? We do not think it can.334

Of course this statement does not show that a conceptual connection between laws and cause could not be established. It merely describes Armstrong’s and Heathcote’s intuitions about the matter. Yet, there is something behind this intuition. Can conceptual analysis provide truths about the world, or does it only tell us about the use of concepts? This was questioned in chapter one when discussing Davidson’s conception of philosophy. Most contemporary philosophers are doubtful that such truths could be established.

I think that against Armstrong and Heathcote, Davidson could nevertheless argue that imagining a world in which causes of the same sort do not give rise to the same sort of effect does not help if we want to consider the nature of causation in this world. The concepts of cause and law under discussion apply in this world, and the question whether a world where

causes are not backed up by strict laws is possible, is not relevant for the analysis of our concepts and thus not relevant for our world. However, since Armstrong and Heathcote argue that the truth of cause-law thesis is empirical, there nevertheless remains the question of how, on conceptual grounds, the necessary connection between a singular cause and underlying law can be proved. According to Davidson, the connection between cause and law is simply due to the fact that “Our concept of a physical object is the concept of an object whose changes are governed by laws.” This being the case, it follows that “The ground floor connection of causality with regularity is not made by experience, but is built into the idea of objects whose changes are causally tied to other changes.” These formulations refer to the important suggestion which Davidson put forward at the very beginning of his career, namely:

I believe the concepts of physical object, of physical event as a change in a physical object or physical situation, and the concepts of physical law are completely tied up with one another. It would be quite misleading to say, first we know what physical objects are, and then we notice what changes they undergo, and then we discover, if we can, whether these changes are governed by laws. I would say rather a physical object is an object whose changes are governed by laws and this is something which is a priori, as Kant said. After all we must mean something by calling something a physical object and I would say this is a very rough sketch of part of what we do mean by calling something a physical object.

Two points are of interest. First, Davidson’s use of the notions “object” or “change” or “situation” refers to the notions as they are being used in everyday contexts. The cause-law thesis is thus grounded in conceptual investigations the target of which is the language of everyday. Second, here Davidson again notes how his understanding of a physical object is a priori, and this in turn is connected to the claim that we must mean something by the concept “physical object” – and what we actually mean by it, in the language of everyday, can be analyzed or clarified. This clearly connects to Davidson’s view according to which something may be “a priori” to us because it is we who decide how concepts are used.

Anscombe and others who wondered what Davidson’s reason for cause-law thesis was would have benefited from reading the lines from 1963, because “Laws and Cause” does not

---

335 We should nevertheless note van Gulick’s observation that situations in which C produces E but C* fails to produce E may be actual according to the current physics which applies to our world.
336 Davidson, 1995c, 214, my emphasis.
337 Davidson, 1985, 227. I take that “idea of objects” refers to our conception of a certain kind of object, to a certain view about the nature of certain kinds of object.
339 For Davidson’s reference to “a priori” see Davidson, 2001d, 1999a. One should note that the claim about the conceptual “a priori” connection between physical object and law was made already in 1963, which suggests that Davidson’s view about the “a priori” and how it relates to his understanding of philosophy was in place from very early on.
add much to the point which was already in place very early. Ramberg’s interpretation, which Davidson applauded, of the reason for holding the cause-law thesis is that we, as biological creatures, have an interest in construing our environment in terms of generalities. The capacity to observe regularities is one that we possess, and what kind of regularities we detect is a result of our biological nature. We cannot help but see the world in terms of regularities. This could perhaps be seen as a naturalistic suggestion of why the cause-law thesis “holds”; it is rooted in our biological nature. Ramberg claims that his point backs up the thesis and concludes:

> When we take two events as causally related, we thereby take them as nomologically related, because to recognize a change in the state of a physical object just is to recognize an event which is susceptible to explanation in terms of empirical laws.\(^{340}\)

Now, whereas Davidson may be correct in his view that our concept of a physical object is an object whose changes are governed by laws and that this conception has its roots in our biological nature, it is mysterious how this could have any relevance if the goal is to give a general argument for monism. Let us suppose, for the sake of argument, that consciousness is a non-physical phenomenon in the sense that consciousness does not occur in space and that it does not obey the laws of physics. If this sounds too wild, one could instead suppose that conscious states are states of the brain, but that they are very strange kinds of physical states because they do not obey the laws of nature, i.e. the fundamental laws of physics, and they thus cannot be described in terms of physics. The brain as described by Freeman would perhaps be the bearer of these kinds of queer physical states. Despite the strange nature of brain states, let us also suppose that consciousness is causally connected to the physical world. Is there anything in Davidson’s argument for monism which would actually challenge the possibility of this kind of dualism? It is hard to see how the argument for AM could show the truth of physicalistic monism. If the reason to hold the cause-law thesis is the claimed conceptual connection between laws and cause, it is difficult to see how this could have any relevance for the question of whether there actually could be causal interactions between physical and non-physical events. As Ramberg points out, it is due to our human nature that we observe regularities in nature and thus, we think, causal connections. But what is the reason to think that singular causal claims and the regularities that we infer from them are supported by the strict laws of physics? Davidson’s claim is that there is a “conviction” that singular causal claims can be sharpened, so that a perfect explanation will eventually be

\(^{340}\) Ramberg, 1999, 611.
reached. The explanation would be perfect in the sense that all the laws to which the explanation refers will be strict. This claim cannot be a conceptual truth, because when one starts referring to actual physics the question what the nature of reality at the bottom level is seems clearly to be an empirical one. The “conviction” is nothing more than a dogma of physicalism, and as such it certainly cannot be used to establish the truth of physicalistic monism. Is it rational to hold this kind of conviction? If there is a perfect physical theory, it is possible that we will never discover it. What is potentially more fatal is, of course, the simple possibility that nature is not deterministic. If there are truly random events, then we can never come up with a theory which could predict them, even if we knew everything about the history of the universe and every physical detail about the nature of reality there was to know.

What justifies the step from the conceptual connection between physical objects, changes and physical laws to the conclusion that causal relations between token-events are such that they can be brought under the strict laws of physics? Whereas the cause-law thesis is a priori and conceptual, Davidson’s views about the status and scope of physics are influenced by empirical evidence, that is, by evidence which is a result of actual research in physics. Although Davidson is hesitant to analyze causation as such, he notes that “[…] if causal relations are ‘in nature’, it makes no sense to classify them as logical or contingent”.341 But if causal relations are “in nature” then the questions of under what kind of laws they fall, how deterministically they can be predicted or explained, and so on, cannot be decided a priori,342 The original formulation of the thesis was: events related as cause and effect fall under strict deterministic laws. But how could it be decided, on a priori grounds, whether or not cause and effect fall under strict deterministic laws? If this cannot be decided a priori then the claim that events related as cause and effect fall under strict deterministic laws is empirical. If it is empirical, then the evidence for monism is empirical. The possibility of ontological dualism cannot be denied a priori. But this would then mean that Davidson’s argument for monism, his theory of mind, would become empirical. Then it would be legitimate to criticize it on empirical grounds.

If the conceptual argument does not give a reason to think that causal relations are backed up by strict laws of physics, then what does? Here Davidson’s somewhat constructive view about physics needs to be noted. He claims, in a Quinean spirit, that it is the aim of physics to find a vocabulary which is complete in the sense that whatever can be described in

341 Davidson, 1985, 224.
342 Indeed, Davidson (1999p, 640) himself notes that strict laws are “as exceptionless as nature permits.” This suggests that the question how strict the basic laws turn out to be depends on the nature of reality. But if this is the case, in what sense is the cause-law thesis a priori?
that vocabulary can also be explained in that vocabulary. According to Davidson, physics is in this sense “all-encompassing”. This resembles the view of Quine who also claimed that: “Full coverage… is the very business of physics, and only of physics.” But even if physics aims at this kind of coverage, there remains the question – and this is obviously the essential point – of whether the nature of reality can be described in the language of physics. Let us suppose that it could never be explained in physical terms why a conscious thought arises in the brain; there seems to be no physical cause for it whatsoever. This is certainly imaginable; it is not a conceptual impossibility. Or let us suppose that we encounter some other events in nature for which no physical cause can be found. How should these situations be understood from the perspective of physicalism? There would be an event, a conscious thought, the occurrence of which would be inexplicable – and thus physics would not have covered reality in full. Davidson himself notes that what the Nomological Character of Causality says is that: “[…]events correctly said to be related as cause and effect would be seen to be covered by the laws of physics if the cause and effect were identified in the vocabulary of physics.”

But what argument shows that all causes and effects can be identified in the vocabulary of physics? One way in which this problem can be solved is just by defining “physical” in terms of physics. This has the consequence that there is nothing physical which could lie outside the reach of physics because, by definition, what cannot be identified in the vocabulary of physics is not physical. But, once again, how could this definitional trick convince us of the view that everything is physical? If a conscious thought remains unexplainable and unidentifiable in the vocabulary of physics, do we therefore accept that it is something non-physical? If so, then the point of dualism is made. It would be absurd to say, as naturalists would, that conscious thought becomes non-existent at the moment its nature becomes scientifically unexplainable.

According to Davidson, the vocabulary of physics is the only vocabulary in which strict laws can be expressed, because physics treats reality as a closed system. But the crucial question for physicalistic monism is precisely whether or not reality is causally closed, in the sense that all physical events have physical causes. If a physical cause for a conscious thought cannot be found, then the thesis of the causal closure of the physical is incorrect – at least if there are reasons to believe that such a cause can never be found. Davidson claims: “It is plausible that there is a set of concepts which lend themselves to the formulation of a closed causal system.”

---

343 Quine, 1981, 98.
344 Davidson, 1999p, 640, my emphasis.
345 Davidson, 1995c, 204.
concepts is “plausible”? How should the term “plausible” be understood in this context at all? Either the physical realm is causally closed or it is not. If Davidson is saying that according to current empirical evidence it seems that physical realm is causally closed, and that the view is therefore ”plausible”, then he should have said that. But he does not refer to empirical evidence at all. What is the reason to believe that when events are truly related as cause and effect, then: “[...] there exists a closed and deterministic system of laws into which these events, when appropriately described, fit.”\textsuperscript{346} This certainly cannot be a consequence of the cause-law thesis as it is explicated in “Laws and Cause”. This cannot be a consequence of any conceptual argument. Why accept the view that: “There is no event that is not fully determined by what goes before, and no state of the universe that does not fully determine what follows... every natural event can be fully explained by the laws of nature and any total prior state of the universe.”\textsuperscript{347} This is one way to formulate the causal closure of the physical order. Instead of claiming that everything physical must have a physical cause, one could claim that any event can be explained. But why believe this either? Are there good reasons to believe, as Davidson does, that: “everything in the universe and its history can in principle be described in the language of physics...”?\textsuperscript{348} All these views are no more than dogmas of physicalism. If they are meant as empirical claims, Davidson should have made this clear. If it should be obvious to the reader that these claims are meant as empirical suggestions, then one is left wondering whether the claims are empirically correct. As has been noted, Papineau’s claim is that philosophers started to believe in the causal closure of the physical because there was empirical evidence in its favor. Davidson, however, does not ever refer to such evidence. He does not refer to evidence which would back up his view. On the other hand, if the claim is empirical, then one is left wondering what its relevance for a priori monism is. The point that needs to be emphasized vigorously is that the philosophical relevance of Davidson’s argument is lost if the claim that everything in the universe and in its history can be described in the language of physics is accepted at face value. One way in which the argument could be understood is that if the nomological character of causality is accepted, then monism follows. This, however, would beg the question for physicalism, which is meant to be the philosophical and metaphysical conclusion of the argument.

\textsuperscript{346} Davidson, 1974b, 231.  
\textsuperscript{347} Davidson, 1993f, 298. Davidson is here describing the view of Spinoza, but the context makes it clear that he could be as well expressing his own view.  
\textsuperscript{348} Davidson, 1997a, 127.
In one of his early articles, Davidson mentions that ontological monism is a view towards which he “announces his adherence” although he does not argue for it.\(^{349}\) In retrospect, it is possible to see that in “Mental Events” the argument for monism was given, and as Davidson notes, the argument for monism was a result of his believing both to the truth of determinism and to the truth of the irreducibility of the mental.\(^{350}\) But, as for example von Wright reminds, if one “accepts pure chance in nature one cannot at the same time postulate strict determinism.”\(^{351}\) Von Wright refers to the work of physicist Ilya Prigogine, who came to the conclusion that the more we know about the universe, the more difficult it is to believe in determinism.\(^{352}\) It remains a mystery what Davidson’s reasons for believing in determinism were since the quoted views – according to which all events are strictly determined and can be fully explained – do not show anything more than the acceptance of physicalistic principles. These kinds of principles were expressed already in the views of the materialists of the previous centuries. Whether they reflect the current scientific understanding of the nature of the reality is doubtful. They nevertheless show Davidson’s commitment to the view that the physical causal realm is closed, which is the cherished dogma of modern physicalism.

The problem is that if one assumes this, then there is no need for any argument to establish monism, given the acceptance of premise one in the argument for AM. If the completeness of physics is true and if mental phenomena causally interact with physical phenomena, this can be only because mental phenomena are physical phenomena. The thesis of the completeness of physics is often used to ground other physicalistic intuitions as well. Papineau argues that supervenience follows from the completeness of physics.\(^ {353}\) Kim uses this principle as an argument for physicalism and for the claim that mental causation is possible only if mental phenomena are physical phenomena.\(^ {354}\) This thesis alone thus carries great weight in the argumentation for physicalism. If Davidson accepts it without argument, then what does he need a philosophical argument for? Davidson himself has claimed that monism does not follow from the nomological character of causality alone, since it is only when this premise is combined with premises with which it is not usually combined that monism follows.\(^ {355}\) I think, however, that Michael V. Antony has convincingly argued that Davidson is able to establish the truth of monism easily without the anomalism of the mental.

---

\(^{349}\) See Davidson, 1963.  
\(^{350}\) See Gluer, 1995.  
\(^{351}\) Von Wright, 1998, 150.  
\(^{352}\) Prigogine, 1997.  
\(^{353}\) Papineau, 1990.  
\(^{354}\) Kim, 2005.  
\(^{355}\) That is, with the principle of causal interaction and especially with the principle of the anomalism of the mental.
– that is, i.e. without premise three which was meant to be the distinctive contribution to the discussion in the original argument.

Let us consider Antony’s claim and suppose that a mental event M causes a physical event P. In the following claim put forward by Antony, P1, P2, and P3 refer to the original premises given in “Mental Events”. P4 is the thesis according to which there are strict laws only in physics. We should note that, as Antony claims:

[…] if all strict laws are physical, then since all causal interactions require subsumption under strict law (by P2), it follows that all causal interactions, including m’s causing p, involve only events that are physical. So m is physical. We have reached monism from P4 and P2! Such an argument, however, leaves no work for the anomalism of the mental (P3) in deriving the identity theory, which is proved before P3 is even mentioned.356

Although Davidson does not explicitly argue for P4, he clearly seems to accept it.357 It is a view which Davidson has defended on independent grounds; yet as we have seen, his reasons for holding it have remained obscure. The point of Antony’s criticism is that if Davidson accepts P4, then taking it together with P2 monism is already established. Then the principle of the anomalism of the mental is unnecessary to establish monism. Kim has earlier made the same observation.358 Worse still, a second way to establish monism without P3 is to assume the causal closure of the physical. As Antony notes:

[…] if Davidson assumes closure, notice how quickly monism follows. All he needs is P1, the premise that at least some mental events interact causally with physical events. For suppose m, a mental event, caused p, a physical event. By closure, all events with which p causally interacts are physical. So m is physical, and we have reached the token identity theory without P3 (or P2 for that matter, if closure turns out to be independent of P2).359

Antony seems to be correct in all of his main points. First, Davidson can establish monism easily given P4 and P2. Second, P3 is thus unnecessary in his defense of physicalism. But the novelty of the argument was meant to be precisely the claim that P3 is required for the establishment of monism. A further problem is that monism seems to follow straightforwardly from the acceptance of causal closure. This, I would say, is a problem for physicalists generally. If cause and effect must be related by a strict law and if all strict laws are physical, it follows that cause and effect must have physical descriptions which satisfy physical laws.

358 See Kim, 1998.
There are good reasons to think that Davidson’s argument for monism is based on assumptions which already show a strong commitment to a broadly physicalistic world view and which remain without convincing arguments. This creates the question of what the whole point of the argument for anomalous monism is. Kim has noted that the causal closure of the physical “[…] isn’t an empirical issue for the physicalists. Or, rather, the point should perhaps be put this way: it is not an empirical issue that physicalism involves the causal closure of the physical domain.” It is no wonder that Davidson’s physicalism involves the idea of the causal closure of the physical, for which he does not give an argument, since once you accept a physicalistic view you accept the causal closure of the physical as well. If the view that physics is able to describe everything is accepted, then the regulative principles of physics may start to seem like a priori features of reality. A problem related to this is that the speculations which Davidson offers about the possibilities of physics may in fact conflict with our current, not to mention future, knowledge about the nature of reality.

As I have noted already, Davidson once admitted that the motivation for the argument for AM was partly based on his belief that “determinism” is true. As we have also seen, in “Mental Events” Davidson’s claim was that there are strict deterministic laws. John Dupre has complained that “the notions of causality and natural law involved in Davidson’s argument are so restrictive that it may legitimately be doubted whether anything whatever satisfies them.” Cartwright and Kim have raised similar worries. Later the requirement of determinism was however dropped by Davidson. We hear: “Since it allows probabilistic laws, the cause-law thesis does not… imply determinism. Neither, then, does it imply complete predictability, even in principle, nor retrodictability.” I have asked what the rationale is for believing that physics would ultimately constitute a closed system. Here we see that Davidson himself admits that physics may actually be incurably non-deterministic and may not deal with closed systems. This view, in contrast to the view that there is a system of strict deterministic laws, is perhaps a result of the progress in quantum physics, of which Davidson became aware. The content of the cause-law thesis is thus subject to the development of physics, and therefore it is unclear how the reason for holding it could be a priori. If Davidson accepts that physics may deal with non-closed systems, then it is unclear how he could answer criticisms like Freeman’s, which was described earlier.

360 In an interview retrieved which can be found at http://www.ephilosopher.com/page.php?15
361 Dupre, 1988, 45.
362 Davidson, 1995c, 205.
Davidson claims that the conceptual connection between the concept of causality and the concept of a law may be tight enough to support completely deterministic laws. But if it turns out that the ultimate laws of the universe are probabilistic, then also causality must be seen as probabilistic. Singular causal statements still entail the existence of “strict” laws, but probabilistic ones instead of deterministic ones. The question about the “amount of determinism” in reality seems to be empirical and, without having any specific knowledge about this, I believe one could argue that current results in quantum physics raise worries for the idea of determinism and for the meaning and nature of “existence” in general. Can Davidson’s conceptual argument ignore the possible consequences of quantum physics? After speculating on whether strict laws should be seen as deterministic or probabilistic, Davidson gives a rough example of a possible strict law: “[…] whenever there is a certain distribution of forces and matter in a field of a certain size at time $t$, it will be followed by a certain distribution of forces and matter in a field of a certain size at time $t’$.” It seems to me that the term “certain” here must mean that if there were to occur twice the same distribution of forces and matter in a field of size $S$, each instance would be followed with identical distribution of forces and matter at time $t’$. I assume that Davidson intends that the expression “certain distribution of forces and matter” can, in theory at least, be replaced by a perfect description, which is formulated in the language of physics. Then the claim would be that from a physical situation $A$ (a certain slice of space time) follows a physical situation $B$ (a different slice of space time), and it would be possible to tell, in physical terms, why this is the case.

Let us suppose that this applies to a case where a mental event $M$ causes a physical event $P$, which can be described also as action. $M$ has been re-described in the language of physics in a way which allows the description of this event as certain distribution of forces and matter at time $t$. In order to explain “the causal nexus” between $M$ and $P$, we have to include everything in the universe “within the sphere defined by the distance light travels in the interval from $t$ to $t’$ from the region of the cause to the region of the action.” From a physical perspective, “the cause” would thus be the totality of distribution of forces and matter inside a field of a certain size. In my opinion all this sounds very mechanistic and it is

---

363 Nevertheless, Davidson speculates that quantum physics may not be the last word and refers to the views of Steven Weinberg (1993) who dreams of a unified theory of everything which would be completely deterministic. Yet, Davidson also notes that the possibility of such a theory is not a purely conceptual question, since there are empirical limitations on the theory that we can come up with due to our nature as human beings. Davidson thus concludes: “Surely we must allow that the best physics that is possible for us is irreducibly probabilistic.” (Davidson, 1995c, 219)

364 Davidson, 1999p, 639.

365 Davidson, 1999p, 639-640. The same point is made in Davidson, 1993d.
not at all certain that every instance of causation could be explained, even in principle, by this kind of model.

Davidson’s views about causality may be at odds with current scientific knowledge about the nature of causality. It needs to be acknowledged that Davidson was hesitant to analyze the nature of causation, although the example above gives a clue about his view. Since we do not know how Davidson saw the current state of physics, there is little need to discuss the problems of quantum mechanics but it should nevertheless be noted that, according to many physicists, the results of current physics may create serious philosophical problems. For example, a fundamental principle of present day quantum mechanics is that the quantum predictions are statistical. John Bell, one of the major figures in 20th century physics wrote that quantum mechanics is simply not able to explain why specific events happen.\footnote{Bell, 1990.} As we have seen, this claim does not count against Davidsonian strict laws, which are allowed to be statistical. But even if it were possible to say, with respect to any event, that the event is going to happen with such and such probability, it is not possible, not even in principle, to arrive at accurate complete and detailed prediction and description of the particular process resulting in a particular event. As the distinguished physicist Anton Zeilinger writes:

I propose that [the] impossibility of describing the random individual process within quantum mechanics in a complete way is a fundamental limitation of the program of modern science for arriving at a description of the world in every detail. In other words, I propose that this is evidence of an element in the description of nature which escapes rational dissection in detail into constituent parts.\footnote{Zeilinger, 1996, 4-5.}

There is thus evidence from scientific research showing, pace Davidson, that it is not the case that the nature of reality can be described in every detail. It also seems that the simplistic compositional picture of reality, dear to many philosophers, is something which does not make sense at the deepest level of reality. In an attempt to find the physical causes of physical phenomena, it is tempting to think that the most fundamental causes are to be found at the most fundamental level. But as another famous theoretical physicist John Archibald Wheeler writes: “Why demand of science a cause when cause there is none?”\footnote{Wheeler, 1983.} According to Wheeler, the individual process in quantum mechanics, the quantum phenomenon, involves an elementary act of creation. This is result of the fact that, at least on some interpretations of quantum physics, the measurement of quantum phenomena affects the phenomena to be measured. According to these interpretations we decide, by choosing the measuring device,
which phenomenon can become reality and which cannot. This should not be understood only as an epistemological claim according to which we cannot know what the nature of reality is before we measure it. The claim is that there is no such reality, that the notion of such a reality does not make sense.

Many physicists have been participating in philosophical speculations when trying to explain the findings of quantum mechanics. What the possible consequences, if any, of quantum mechanics will be for the mind–body problem remains to be seen, although the philosophical discussion about the relation of quantum mechanics and consciousness or free will is already active among philosophers. Quine claimed that: “Quantum mechanics today, indeed, in its neoclassical or Copenhagen interpretation, has a distinctly mentalistic ring”\(^{369}\) and suggested that current understanding of physics questions the meaningfulness of the question of “what there is.” Quine’s view is in conflict with his earlier position, which was accepted also by Davidson. According to this position, physics describes the world “as it is”. Perhaps there will come a time when the view that science can describe the true nature of reality must be abandoned also by philosophers. Zeilinger, describing the conclusions of his research, writes:

> It’s all pretty crazy. The spooky effect at a distance is a process outside time and space that even I can’t really imagine. But I believe that quantum physics tells us something very profound about the world. And that is that the world is not the way it is independently of us. That the characteristics of the world are to a certain extent dependent on us.\(^{370}\)

Given the debate about metaphysical realism among philosophers, it is interesting that many of those philosophers who vigorously defend realism, because of its “scientific plausibility”, ignore the views of those scientists who actually study the most fundamental nature of reality. When the empirical evidence is applied to Davidson’s view about causal relations in nature, it simply may turn out that there are no such relations.

As far as Davidson’s argument for physicalistic monism is concerned, I think we have to conclude that there is no way that it could succeed without being circular and relying on broadly physicalistic assumptions. Davidson’s argument allows an insufficient amount of monism. First of all, as Antony notes, the possibility of establishing monism without the argument structure for AM seems clear. It can be admitted that part of the novelty of Davidson’s argument still stands. His purpose is to show that mental events are physical events, even though they cannot be reduced to them. If the anomalism of the mental stands, it

---

\(^{369}\) Quine, 1995, 257.

\(^{370}\) Zeilinger, 2006, 5.
is shown that ontological reduction does not imply definitional reduction. Moreover, one may have interest in the argument because it tries to show how three principles which appear to be true can be combined despite their apparent inconsistency. Nevertheless, the argument is not required for monism if the causal closure of the physical is assumed. Second, the explication of the cause-law thesis which is given in “Laws and Cause” remains obscure and the question, posed by many critics, of why the thesis should be accepted still remains without a clear answer. I believe it is legitimate to say that Davidson never gave a convincing reason why the cause-law thesis should be accepted. The conceptual argument which he gives in “Laws and Cause” does not explain why causal relations in nature would be such that they entail the existence of strict laws. Third, it seems that Davidson’s view of physics does not fit well with the actual state of modern physics. All the speculation about the full coverage of physics and the vision of a closed deterministic system remains without good arguments. Actual physicists see the task of physics quite differently. It is thus easy to agree with Fodor when he says:

..I don’t pretend to do what Davidson seems to think he can, viz., get physicalism just from considerations about the constraints that causation places on covering laws together with the truism that psychological laws aren’t strict. That project was breathtakingly ambitious but maybe not breathtakingly well advised. My guess is, if you want to get a lot of physicalism out, you’re going to have to put a lot of physicalism in.  

Davidson’s “breathtakingly ambitious” project to give a philosophical argument for physicalism stands or falls with the claim that causation must be backed up by strict laws and the problems of this view are evident.

As I have emphasized, Davidson’s argument for monism differs from the arguments of the type–type identity theorists in the important sense that whereas the latter offered monism as a scientific hypothesis, Davidson wanted to give monism an a priori status. This being the case, I think that those critics accusing Davidson of half-hearted physicalism neglect that even an attempt to give an a priori argument for physicalism shows a radical bias towards a certain kind of view about the nature of reality. Davidson’s argument would be an important contribution to physicalism – if it succeeded. Those interpretations which emphasize that Davidson can be seen as an eliminativist about the mental, sharing in essential respects Quine’s physicalistic intuitions, are not completely mistaken. Although Davidson’s reasons to argue for the anomalism of the mental are “Kantian”, in the sense that he wants to show how human freedom is possible, I suggest that Davidson’s conclusion is actually the mirror image of Kant’s conclusion. Davidson acknowledges, as his defense of the anomalism of the mental  

371 Fodor, 1989, 79, my emphasis.
clearly shows, that there are important differences between the mental and physical domains, but he also shows his special allegiance to one ontology, that of natural science, or more precisely, that of physics. By doing this he, in my opinion, unfortunately retains some of Quine’s scientism, which is expressed also in the views of contemporary naturalists.

On the Davidsonian–Quinean view, the autonomy of the mental must be sought within the constraints of a physicalistic ontology. It could be said, tongue in cheek, that whereas Kant assumed freedom and tried to defend the autonomy of the mental within the framework of experience, Davidson assumes determinism without much argument and tries to defend the autonomy of the mental in the framework of physicalism. The starting point is a commitment to the ontology of physicalism, and the subsequent attempt is to place mental phenomena within this ontology. In this sense Davidson, after all, quite simply follows the naturalistic consensus according to which something must be said about the mind’s place in a physical world. I have to conclude that there is thus an intrinsic conflict in Davidson’s views. On the one hand, as I have argued, his position could be interpreted as coming close to classical naturalism, which does not subscribe to the view that the ontology of physics (or any ontology) would be in any sense privileged. On the other hand, Davidson’s position shows a bias toward a strict physicalistic ontology. The threat which mental phenomena could pose for the completeness of physical explanations is resolved by insisting that mental phenomena are physical. Although Davidson’s argument for physicalism fails, it can be concluded, in Skillen’s words, that Davidson is “an empirical idealist but a transcendental physicalist.” I agree with this definition while noting that there are reservations which I will discuss in section 2.6.

I have described the general nature of Davidson’s physicalism and the insuperable problems facing it. The argument structure of AM is meant to be a strong argument for physicalism, but the physicalistic content of Davidson’s non-reductive physicalism is vague and the argument itself is problematic. In the next section, I consider von Wright’s physicalistic view in order to find out whether a general argument for physicalism and a plausible version of non-reductive physicalism can be found within it. Both von Wright and Davidson accept physicalistic monism, but their respective reasons for accepting it are interestingly different.

372 Skillen, 1984, 523.
2.2 Von Wright’s physicalism and its relation to Davidson’s position

In this section I clarify von Wright’s views and discuss the sense in which von Wright can be interpreted as a physicalist. I emphasize those aspects of von Wright’s views which bear similarities to Davidson’s. The connection is interesting given the fact that Davidson and von Wright basically never mention each other; yet they have reached similar kind of conclusions, as far as we can tell, on independent grounds. It is also interesting that commentators have interpreted their views as being diametrically opposed on many issues about the mind.

Classical philosophical questions about the mind stimulated von Wright’s interest in philosophy already in his early youth. At the end of his career von, Wright became preoccupied with the mind–body problem. The result was his book *In the Shadow of Descartes*, containing essays dealing with several problems in the philosophy of mind. Apparently, the book contains texts which von Wright did not originally plan to publish.\(^{373}\) It is fortunate that the book was published; I agree with Hertzberg that the work contains some of von Wright’s deepest ideas.

Von Wright did not write much about the mind–body problem or about the questions which fall under the label “philosophy of mind” through most of his career.\(^{374}\) Consequently, his view has received little attention or discussion, and therefore von Wright’s position in the mind–body debate does not enjoy anything like the established status of Davidson’s AM. There are literally only a handful of subsequent commentaries on von Wright’s work in the philosophy of mind.\(^{375}\) Von Wright is not putting forward a detailed theory of the mind, and in this sense his contribution to the contemporary discussion is somewhat feeble. The insights of Wittgenstein’s philosophy of mind or psychology are largely ignored in contemporary philosophy, and the same can be said in von Wright’s case. An interesting question is that of whether the fact that von Wright’s views go largely unnoticed has anything to do with the fact that in the eyes of many he is a “Wittgensteinian.” I shall not speculate about this, although it can be noted that modern philosophy of mind seems to be hostile towards a “Wittgensteinian” point of view. What are the other possible reasons why von Wright has stayed outside the contemporary discussion, i.e. why has he remained in the “shadows” of modern philosophy of

---

\(^{373}\) As told by von Wright to Hetzberg according to the latter. See, Hertzberg, 2007.

\(^{374}\) In addition to few articles on the subject, von Wright’s views about the mind are presented solely in *In the Shadow of Descartes*.

\(^{375}\) Stoutland is the most active philosopher commenting on the work of von Wright, but also his analyses of von Wright’s philosophy of mind are few. The best commentaries on von Wright’s work later work can be found in Meggle (ed.), 1999, Egidi (ed.), 1999 and Stoutland (ed.), 2009. In addition to these interesting commentaries are Jacob, 2002 and Egidi, 2009, Emiliani, 2001, 2005, UP. Von Wright is cursorily mentioned by Svensson, 1994 and Lagerspetz, 2002.
mind? One possible reason is that, as von Wright himself noted, the direction in which modern philosophy was heading was not to his taste.\textsuperscript{376} Staying true to his understanding of what philosophy is and believing that there are no final solutions to philosophical problems made von Wright feel that: “Many things which philosophers say, naively simplifying matters, I could never say…”\textsuperscript{377} It was the unhealthy materialist trend of modern philosophy of mind which led von Wright to keep a distance between his views and the current debate. Although von Wright described eliminative materialism or physicalistic reductionism as the most vulgar examples of the scientistic tendencies in the modern philosophy of mind, he did not challenge these or other positions actively in his writings.\textsuperscript{378} This I take to be unfortunate; perhaps it could be counted among the responsibilities of a great philosopher that he should comment on the trends of time which he sees as troubling. Bennett and Hacker’s invocation of Wittgensteinian insights against the conceptual confusions of neuroscience provides an example of contemporary philosophers of mind carrying on this kind of project. But the requirement to participate cannot be forced on anyone, and perhaps von Wright felt too alienated from the mood of our times to participate; he noted that even reading the newspapers was painful because they reminded him of the world’s suffering and contained so much miserable material. On the other hand, von Wright actively participated in public discussions concerning societal issues, and these contributions ranked certainly higher than the public comments of most philosophers of his home country. I think it is clear that In the Shadow of Descartes contains a “wealth of distinctions and observations which may prove productive”, as Stoutland has suggested.\textsuperscript{379} A Finnish commentator, Jarkko Tontti, has claimed that von Wright’s production is not relevant any longer and that, truth be told, Explanation and Understanding is a completely insignificant contribution to the current philosophical discussion.\textsuperscript{380} By exploring further the views of von Wright, my purpose is to show that, pace Tontti’s sensationalist claim, von Wright’s work is still relevant and an important contribution to the philosophy of mind.

Von Wright’s views are problematic if seen in the context of naturalism. Those who expect a technical view about the relationship between mind and brain will certainly be disappointed by von Wright’s views; given his conception of philosophy this was, of course, to be expected. But von Wright’s remarks on the problems which are being currently

\textsuperscript{376} In an interview by Lahtinen, Pihlström and Tuusvuori, 1995.
\textsuperscript{377} Von Wright, 1989a, 16.
\textsuperscript{378} See von Wright, 1995.
\textsuperscript{379} Stoutland, 2009, 9.
\textsuperscript{380} Tontti, 2004.
discussed in the philosophy of mind serve as a valuable starting point from which the consistency and possible problems of the dominant naturalistic views can be considered. Von Wright’s position is an interesting alternative, and its merits and defects deserve a discussion. His views are fresh and invite the reader to think by offering suggestions instead of compelling knock-down arguments. When considering the relevance of von Wright’s views on the mind–body problem, his views about the nature of philosophy (as described section 1.2.3) must be taken into account; the discussion must be understood against this background. Von Wright does not think that the task of a philosopher is to come up with the kinds of explanatory theories which are now so common in the naturalistic philosophy of mind. Yet, as I have noted, von Wright is not always true to his conception of philosophy and – when this happens – his views face serious empirical challenges.

Von Wright thinks that the mind–body problem is, above else, a problem which, instead of an empirical solution, is in need of conceptual clarification. As he writes: “[…] no future ‘brain research’ will – contrary to what many enthusiasts now seem to expect – eventually solve ‘the riddle of consciousness’. This… is a philosophic muddle which no findings of a scientific nature will ever clarify.”. This kind of very provocative claim sounds astonishing, or simply false, to those scientists and naturalistic philosophers who see the mind–body problem as nothing other than an empirical issue. Searle, for example, with his biological naturalism, would see this kind of approach as utterly wrong-headed. Those, like Nagel and McGinn, who see the mystery of consciousness as more of a philosophical problem would perhaps have more sympathy for von Wright’s views. I think it is important to acknowledge that von Wright has his own reasons for making this challenging claim, reasons which result from his conception of philosophy.

Von Wright’s views are sometimes obscure. The presentation of In the Shadow of Descartes shows, in my opinion, how the author tries to clarify the issues, above else, to himself. This is not surprising because, as I described in chapter one, von Wright’s view is that in the end a philosopher can appeal only to his own intuitions when seeking relief from a philosophical problem. Whereas easily-followed arguments or conclusions are sometimes lacking in the views of von Wright, deep philosophical insights are always present. In section 2.1, Davidson’s physicalistic monism as an answer to the ontological mind–body problem was discussed. An acceptance of physicalistic monism can be detected also in the writings of von Wright. He, in a similar spirit as Davidson, would suggest physicalism as an answer to

381 Von Wright, 1989b, 30.
the mind–body problem. A critic could claim that (because of their non-reductive views) neither Davidson nor von Wright is happily seen as a pure physicalist, and that they subscribe to a form of monism of a non-physicalistic variety. As will become clear, there are reasons to think that the position of these philosophers could be described also in this way. Nevertheless, both Davidson and von Wright reject ontological dualism and are, in this sense at least, physicalists and do not belong to the group of the (few) modern dualists. Although von Wright’s position is not a form of dualism, it is true that his work is alien to the kinds of positions which are usually identified being part of the physicalistic ideology.

I suppose that von Wright saw most of the contemporary work in the philosophy of mind as being very different from the kind of conceptual considerations which interested him. He thought that the materialistic spirit of modern philosophy of mind was unhealthy and that this kind of philosophy contained some grave errors. This being said, von Wright’s position can justifiably be described as a form of physicalism, and in my opinion it is clear that he, like Davidson, fails to give a convincing argument for its support. My claim is that insofar as von Wright’s view is interpreted as a form of physicalism, his position clearly resembles Davidson’s position. According to my interpretation, the similarity of von Wright’s and Davidson’s versions of physicalism casts doubt on the claim, made by certain commentators, that an important difference between them is that the latter endorses the idea of supervenience whereas the former does not. This claim, which I believe to be incorrect, makes Davidson “more of a materialist” than von Wright, according to the critics. However, despite the similarities I argue that von Wright’s and Davidson’s arguments for physicalism differ in an important way.

Both von Wright and Davidson are ontological monists. In the previous section it was argued that Davidson’s position is a form of monism, and the question which required clarification was what sort of monism it is. So far my interpretation has strongly emphasized the physicalistic aspect of Davidson’s monism. What remains to be considered in the next sections are the questions of what this monism implies about the nature of properties and their reducibility, about the nature of token-states or events and their reducibility, and about the relation between mentalistic and physicalistic explanations. The answers to these questions reveal the exact nature of Davidsonian non-reductive physicalism. Although non-reductive in spirit, the physicalistic monism of von Wright and Davidson is committed at least to this: “[…] destroy the physicalistic and you destroy everything else though the contrary is not

\[382\] See the interview by Lahtinen, Pihlström and Tuusvuori, 2005.
\[383\] This view is put forward, for example, in Emiliani, 2001.
This, according to Stoutland, is one way to express the idea of the global supervenience of the mental on the physical. Stoutland, who knows the work of Davidson and von Wright very well, would say that both of these philosophers are best understood as non-physicalists. This, I think, is because according to him almost any philosopher, apart from the “supernaturalists”, would subscribe to minimal physicalism (which is equivalent to what I have labeled O-physicalism). The fact that a philosopher accepts minimal physicalism is not, according to Stoutland, a sufficient reason to label him as a physicalist. The question of when a position is physicalistic enough to be called “physicalism” is controversial, and is partly a terminological question. Kim, for example, claims that mind–body supervenience, the anti-Cartesian principle and mind–body dependence define minimal physicalism, and that if one accepts them all then one may be properly called a physicalist. Consequently, a rejection of one or all of them means the rejection of physicalism. As far as I can see, both von Wright and Davidson accept the three theses which capture the essence of minimal physicalism according to Kim. Both Stoutland and Kim would thus describe Davidson and von Wright as minimal physicalists, but both would also claim that this is not “physicalism enough”. Coming from Kim this is criticism, whereas I suspect that Stoutland would not see it as a defect but a virtue of von Wright’s or Davidson’s position.

With respect to Davidson, Stoutland notes that there is no question that Davidson is a physicalist in the minimalist sense, but he also rejects all the claims of robust physicalism. Stoutland concludes that Davidson’s philosophy of mind goes, in “all major respects”, against robust physicalism. This is reminiscent of Rorty’s interpretation of Davidson as a pragmatist showing no interest in a physicalistic ontology. But we have seen that Davidson’s statements about the task and possibilities of physics make it far from certain whether it is correct to claim that his view goes, “in all major respects” against robust

---

384 Stoutland, 2005, 147.
385 At least this is the feeling that I have from a personal discussion. According to Stoutland, both von Wright and Davidson have a monistic view which is not physicalistic.
386 Kim, 1996, 12. Mind–body supervenience says that two things exactly alike in all physical properties cannot differ in respect of mental properties. The anti-Cartesian principle says that there can be no purely mental beings. Mind–body dependence says that mental properties are determined by physical properties.
387 Stoutland, 2008. Robust physicalism I take to be something like M-physicalism.
388 I am not suggesting that Stoutland would agree with Rorty’s view but merely noting that their views about the nature of Davidson’s physicalism seem similar. I believe Rorty’s comment is best understood as a description of the intellectual camp in which he takes Davidson as belonging. That is, Davidson is a pragmatist whose aims in philosophy differ from those of contemporary physicalists. Stoutland would quite likely agree with the view that Davidson understands the nature and purpose of philosophy in a different way than contemporary physicalists. This is, of course, the point I wanted to stress in section 1.2.2.
physicalism. It certainly does not go against robust physicalism as radically as openly
dualistic views, and Davidson’s claim that physics is capable of giving full coverage and
perfect explanations does not sound very anti-physicalistic. The question that I want to raise is
precisely that of what the reasons to accept minimal physicalism ultimately are, and more
specifically, what the reasons that Davidson and von Wright give for minimal physicalism
are. Does von Wright offer better reasons for physicalism than Davidson?

“Minimal physicalism” or not, non-reductive physicalism is nevertheless physicalism,
at least by name, and the pressing question is whether even the minimal physicalism of this
position is coherent and justified. To say, as Stoutland does, that “almost any philosopher”
would accept minimal physicalism obviously does explain why a certain philosopher accepts
such a position or why the position of minimal physicalism should be accepted. Whether this
claim is true is worth considering. Ignoring the vague expression “almost any philosopher”,
the claim is false since there are quite a few modern dualists as well as philosophers with
religious backgrounds. All these philosophers with different arguments can be brought
under the (possibly) somewhat pejorative label “supernaturalists”, but doing this does not
clarify the question of whether physicalism or dualism is the correct answer to the mind–body
problem. Why is it that certain philosophers reject physicalism?

It could be debated whether von Wright would have accepted the term physicalism as
a correct description of his position, but it is clear, as I will show, that the idea of global
supervenience expresses well von Wright’s view about the overall relationship of mind and
matter. Von Wright uses the concept of supervenience also in a more substantial sense. This
being the case, the following interpretation by Emiliani, who knows von Wright’s work well,
should be treated with caution: “Von Wright’s conception basically departs from Davidson’s
– as well as from several approaches to the mind–body problem focusing upon the notions of
supervenience and emergence….” According to Emiliani, von Wright’s view departs from
these approaches precisely because it does not focus upon the notion of supervenience and / or

---

389 In most respects I agree with Stoutland’s view. My claim is merely that some of Davidson’s own views give
the impression that he would not reject all the claims of robust physicalism. I claim that a Davidsonian
conception of the mind is anti-physicalistic, but it is possible that I am reading more “dualism” in Davidson’s
position than what he himself intended.

390 What is true is that almost any modern philosopher with a naturalistic view accepts minimal physicalism.
With respect to all philosophers one should note, for example, the claim of Chalmers (2005) who writes: “If I
had to guess, I’d guess that the numbers within philosophy of mind are 50% materialist, 25% agnostic, 25%
dualist.” It is questionable whether the physicalistic consensus is as established as Stoutland claims. The view is
challenged by notable philosophers working on the field. Stoutland himself does not offer arguments for minimal
physicalism and merely states that he accepts global supervenience. Perhaps he thinks that the position needs not
to be argued for.

392 I do not mean to suggest that Stoutland uses this expression in a pejorative manner.

emergence. Emiliani also claims that von Wright rejects the view that mental and physical phenomena are token identical, and thus his position is a very weak form of materialism if it can be described as materialism at all. Kim described Davidson’s position as a weak form of token physicalism because while token-identity is argued for, the possibility of actual token-reduction is questioned.\textsuperscript{393} If Kim’s description is correct, it is plausible that Emiliani fails to see the similarity between Davidson and von Wright because, on the one hand, he interprets Davidson’s views as implying a strong form of token physicalism, which includes the idea that tokens of mental events could be actually reduced to tokens of physical events and, on the other hand, he thinks that von Wright rejects the token-identity thesis altogether. As the different interpretations of Davidson’s position show, the strength of Davidson’s physicalism is not easily determined, and therefore Emiliani’s interpretation cannot be straightforwardly dismissed. My claim is that with respect to von Wright’s views about token-identity Emiliani’s claim is however incorrect.\textsuperscript{394}

As I understand him, Emiliani thinks that an approach which “focuses” upon the notions of supervenience and emergence is a form of strong materialism, because on this kind of approach the materialist conceptualization constitutes the most basic level of descriptions. What it means to say that a position “focuses” upon the notion of supervenience is in my opinion unclear, as well as the question of how the notion of “basic level” should be exactly understood. When is one description more basic than another? Perhaps one way to clarify the view that descriptions form a hierarchy in which some descriptions are more basic than others is just by invoking the concept of supervenience.\textsuperscript{395} Davidson’s acceptance of supervenience is indeed one important reason to see him as a materialist, because according to him physical facts determine mental facts and the latter are dependent on the former. But Davidson does not put any special emphasis on the concept of supervenience; his view about the relationship between mind and body does not “focus” on the concept of supervenience. If the view focuses on anything, it is on the irreducibility and the anomalism of the mental.\textsuperscript{396} Emiliani argues that on Davidson’s view “matter is ontologically prior to the mind”, and that the existence of physical events is language-independent whereas the existence of mental events is not.\textsuperscript{397} I think that the distinction between language-dependent and language-independent events

\textsuperscript{393} Kim, 2005.
\textsuperscript{394} I return to this question in section 2.4 where the problem of token-identity will be considered.
\textsuperscript{395} Although as we have seen in section 2.1.2, it is questionable whether the notion of supervenience offers any substantial clarification with respect to the mind–body problem.
\textsuperscript{396} Of course, as I have argued, Davidson also focuses on establishing the truth of monism. However, in this line of argumentation the concept of supervenience does not play a significant role either.
\textsuperscript{397} Emiliani, UP. It seems to me that, according to Emiliani, Davidson is an eliminativist claiming that mental phenomena are mere constructs.
describes Davidson’s position rather uneasily, because at times he argues that events are mental or physical only as described. Therefore it is correct to say that the form of supervenience which Davidson seemingly accepts implies no more, and no less, than what Stoutland notes – that is, destroy the physical and you destroy everything else. If this is all that the supervenience thesis should be understood to mean, then, pace Emiliani, both von Wright and Davidson should be seen as holding the thesis of supervenience. If one interprets von Wright as claiming that mental phenomena are not token identical to physical phenomena, then one is apt to see him as a substantial anti-physicalist – even as a dualist – which, so I argue, is an incorrect description of his position.

To say that von Wright accepts a form of materialistic monism does not clarify much if the content of this monism is unclear. How does the irreducibility of the mental, which von Wright surely accepts, go together with materialistic monism? This question, which is a problem for Davidson’s position, is a problem also for von Wright. As the title of von Wright’s last book suggests, his view is that the contemporary philosophy of mind has been pursued “in the shadow of Descartes”. Descartes made a sharp distinction between the mental and physical aspects of reality. Yet, the two realms seem to interact. The now all-too-familiar question, which is asked in the shadow where the mental and the physical have been separated, concerns the relation between the mental and the physical. If the starting point is the intuition that “here” is mind and “there” is matter and that they are intuitively two kinds of different “things”, then the question about their relationship arises. A sharp distinction results in a situation where either a form of interactionism or a form of idealist, materialist, or neutral monism (reductionism) must be accepted. From among these alternatives, it is clearly materialist reductionism which is favored among modern philosophers. The naturalistic attitude in philosophy is a major reason for this. Real alternatives to physicalism or dualism are few.

Von Wright, however, rejects both alternatives as successful answers to the mind–body problem. He claims that both solutions, interactionism and reductionism, belong within the Cartesian frame and thus err right in the beginning by accepting the sharp distinction that Descartes made famous. This criticism bears an interesting resemblance to the view discussed in connection with Searle’s biological naturalism. His claim is that the main mistake of contemporary philosophy of mind is its inability to free itself from the Cartesian grip. Von

398 If this is what Emiliani means by the claim that “matter is ontologically prior to the mind”, then it is unclear why he emphasizes that this is peculiar to Davidson’s position if global supervenience is something that almost any philosopher would accept.
Wright agrees with this claim, and describes his own approach in the philosophy of mind as an independent track in the direction pointed out by Heidegger, Husserl, and Wittgenstein who all attempted to break free from the shadow of Descartes. As a personal voyage, philosophy of mind was for von Wright an attempt to release himself from the Cartesian ghost which has haunted philosophy. It could be claimed, as Searle does, that the philosophical problems of the mind are still formulated in surprisingly Cartesian terms. The formulations mind and body, or mind and brain, are very common. The self-image of the human is still dualistic, even though the scientific image of the human may points towards monism. Von Wright’s claim is that contemporary physicalism in fact maintains a Cartesian-like distinction between an inner mind and the outside world, or between persons and brains. The same claim has been made recently by Bennett and Hacker. Von Wright thinks that the sharp distinction between mind and body should be removed, but at the same time he claims that there is an obvious conceptual difference between mind and matter. An attempt to remove or reject mind–body dualism through reduction or identification would ignore the conceptual subtleties which make up this very dualism. I think it must be noted that von Wright’s final views about the legacy of Descartes are somewhat unclear. On the one hand he notes that in his opinion the body–mind distinction is not, pace Descartes, a sharp divide. On the other hand, in an unpublished manuscript where von Wright states his last remarks on the subject, he writes: “The unquestioned basis of my paper could be called Cartesian. It assumes that there are two kinds of phenomena which are (conceptually) sharply distinguished and irreducible the one to the other.” In another unpublished manuscript von Wright concludes: “Perhaps one could call my position a revindication of some Cartesian intuitions.” These are curious remarks, which may reflect merely the fact that von Wright changed his mind about the question of how Descartes’ challenge or shadow should be faced.

Von Wright’s provocative view is that modern philosophy of mind goes astray in two fundamental aspects. First, it maintains the sharp Cartesian-like separation of mind and world; and second, it is based on the assumption that this separation can be dispelled through reduction. By making these two false assumptions, the program of reductive physicalism is committing a grave mistake and heading in the wrong direction from the beginning. Von

---

400 Searle’s complaint is that too many philosophers still accept the “common-sense self-image” of human beings instead of the scientific image.
403 Von Wright, UPe, 1.
404 Von Wright, UPe, 1.
Wright thought that contemporary philosophy of mind was losing itself to the jungle of metaphysics, which is a view shared among recent anti-naturalists like Putnam and Rorty. But if a reductive view about the mind is fundamentally flawed then what, according to von Wright, are the reasons we might give to reject interactionism? What are the reasons one might have to argue against the view that mental and physical phenomena are ontologically distinct? This question is the reverse side of the question of why materialistic monism should be accepted in the absence of reduction. Philosophers’ reasons to accept materialistic monism are various, but a convincing argument for monism is lacking. Davidson tried to provide such an argument, but failed. Whereas Davidson attempted to come up with a philosophical argument, von Wright does not give this kind of argument for monism. His reason for rejecting interactionism is straightforward and actually reflects quite simply the attitude of naturalism. Von Wright writes: “[I am]… reject[ing] the idea of interactionism as contrary to a scientific picture of the world.”\footnote{Von Wright, 1994, 148.} The view that something material could have an immaterial cause, or that something immaterial could cause something material, goes against our ‘scientific picture of the world’. In a similar way as we observe in the views of many physicalists, the causal argument is used to reject the possibility of dualism in von Wright’s reasoning as well. Von Wright does not discuss the details of why, with respect to causation, the situation should be so.\footnote{Perhaps the empirical evidence for the completeness of physics is strong enough evidence for von Wright. If this is so, he does not mention it.} He merely notes that he cannot understand how something immaterial could cause a material event; for von Wright this puzzlement seems to be both philosophical and scientific. In von Wright’s case, as in the case of most modern philosophers of mind, the problem of mental causation is nevertheless \textit{the} problem which calls for an analysis of the mind–body relation, and which points towards a monistic solution. A further question is arises as to why von Wright is willing to defend a “scientific picture of the world” even though he at the same time is very critical of scientific progress and scientism. The only answer I can propose is that von Wright inherited from his teacher Eino Kaila a deep respect for the results of exact sciences, especially physics, and for the idea that one of the pillars on which monism rests is the unity of the scientific world picture. To this view von Wright certainly subscribed in the beginning of his career.

What I want to emphasize is that that von Wright’s reason for accepting monism differs in an interesting way from Davidson’s. Whereas Davidson wanted to give an \textit{a priori} philosophical argument for physicalism, von Wright seems to take a form of materialist
monism for granted. He notes that interest in reviving a form of interactionism can be detected in the contemporary philosophy of mind but, as far as he can tell, taking the situation back to the “neighborhood of Descartes” can be a result only of accepting a kind of pre-scientific, “animistic” or “magical” way of thinking. This way of thinking should be avoided and opposed. As an example of an “atavistic attitude” towards the mind–body problem, von Wright refers especially to the work of Eccles and Popper who subscribe to a substantial form of dualism which, in my view, certainly cannot be interpreted as a form of physicalism at all.407 An important metaphilosophical question is at what point interactionism became a form of this magical way of thinking. Already Descartes was challenged with the question of how body and mind can interact, but this challenge was not as much scientific as it was philosophical. When did the challenge become a scientific challenge? As I showed in section 2.2.1, the physicalist consensus was not as firmly in place in the 1950s as it is today. The progress of science and the belief in the completeness of physics were among the reasons why philosophers started to believe in the material mind. It would be interesting to consider what kind of evidence von Wright saw as showing that interactionism is unscientific. If he had developed a theory of mind in the 1950s, would he have been together with Smart and others who defended a type–type theory because of its scientific plausibility? Or was it only at the end of the 20th century that interactionism looked implausible from the scientific perspective? On this question I shall not attempt to speculate. It can nevertheless be noted that according to von Wright, it was the growing significance of brain research in the 1970s which partly caused his interest to gradually move towards the philosophy of mind.408

It should also be emphasized that von Wright’s view about the mind–body relationship bears a striking resemblance to Kaila’s view. It seems to be the case that von Wright acquired not only his respect for scientific world view but also his position in the philosophy of mind largely from his teacher. Von Wright notes that for Kaila: “dualism was ‘out’, once and for all. This attitude is, I think, a reflection of the situation in philosophy when he grew up.”409 He goes on to speculate that Kaila would have seen the revival of dualism in the latter part of the 20th century as: “[an] …aberration and relapse into already conquered positions.”410 Von Wright seems to follow Kaila’s path, who in the words of von Wright, “succeeded in showing that the body–mind separation is an unfortunate instance… and that the two Cartesian substances are conceptually inseparably bound together. It does not seem to me certain that

408 Von Wright, 2002, 273.  
409 Von Wright, 1992b, 66.  
410 Ibid.
monistic philosophy can go much further to their unification.” This is an interesting statement because I believe it tells us something also about von Wright’s own position. On the one hand, as we have seen, von Wright thinks that mind and body are conceptually sharply distinguished. This is contrary to Kaila’s view. On the other hand, von Wright accepts Kaila’s claim that mind and body are conceptually inseparably bound together. It should be noted that von Wright has as negative a view of the recent interest in reductive views about the mind as he does towards interactionism. I would therefore claim that von Wright opposes those views which he sees as not scientific enough as well as those which he interprets as being scientistic. In doing this, he tries to steer a middle course; this of course is essentially the position of non-reductive physicalism. A paradox is that the position of von Wright – who accuses interactionists of being pre-scientific – is subject to the same criticism, according to most modern philosophers of mind. This, the “unscientific” nature of von Wright’s work can be counted among reasons why his position has remained in the shadows of modern philosophy of mind.

But the scientific picture of reality has not yet described mind’s place in the physical world. Therefore it is better to say, as von Wright in fact does, that if materialist monism is true, it is contingently true. His view is that an attempt to argue that physicalism is necessarily true would be a hopeless task. Von Wright claims that it is logically possible that there might be mental phenomena without physical correlates, and this possibility is what demolishes the philosophical interest of identity theories. I think most contemporary philosophers of mind would likely agree with the view that materialism is contingently true, although von Wright notes that in modern philosophy the main question about the status of psycho-physical parallelism has been the question of whether such a parallelism is a logical necessity. In this work I shall not take a stance on the question of how the modal status of psycho-physical identity claims should be understood. I acknowledge that the question about the modal status of identity claims has been under debate, and that many see this discussion as an important part of the problem. I choose to largely ignore this discussion in considering the views of Davidson and von Wright. Davidson repudiates all other forms of necessity than the natural kind and has expressed doubts about the idea that identities should be understood

---

411 Von Wright, 1992b, 80.
412 In fact von Wright discusses the necessity of psycho-physical parallelism but given the fact that the essence of this problem is the question of whether there is something physical corresponding to every mental phenomenon, we can conclude that the views about parallelism can be turned into views about the status of physicalism.
413 Modal arguments against materialism are given, for example, by Kripke (1980) and Chalmers (1996).
as being necessary in Kripke’s sense.\textsuperscript{414} I shall not here consider the question of how Davidson understands “natural necessity”; it is an issue which did not play an important part in Davidson’s philosophy, and it certainly did not have an important role in his philosophy of mind. Whereas von Wright considers the modal status of identity theory, it is clear that his view about its truth or falsity is based on empirical evidence. I think that the question of whether mental states are physical states in this world is an interesting and a difficult enough question already. To this issue philosophy can make a contribution, for example, by considering whether there are principled obstacles preventing the possibility of an empirical answer to the question.

Von Wright’s emphasis on the contingency of psycho-physical identity nevertheless differs from Davidson’s view. Davidson’s argument is meant to show that, given certain non-empirical assumptions like the nomological character of causality, the truth of monism is not based on empirical evidence in any clear sense.\textsuperscript{415} Von Wright, on the other hand, thinks that the psycho-physical identity theory is best understood as a \textit{scientific hypothesis and not as a philosophical view}. The question of whether or not mental phenomena are identical with physical phenomena is an empirical and not a philosophical question and, according to von Wright, the empirical evidence seems to point towards the conclusion that mental phenomena are physical phenomena. O-physicalism is true, or there are at least good reasons to believe in its correctness. The \textit{philosophical} question would be whether mental phenomena \textit{have} to be physical, and to this question von Wright’s answer is negative. Davidson, contrary to von Wright, can be interpreted as answering this question affirmatively. Von Wright claims that the interesting \textit{philosophical} question is not whether physicalistic monism is true or not, but whether the idea of monism can be made intelligible.\textsuperscript{416} As I see it, this is due to von Wright’s conception of philosophy and his view concerning the philosopher’s task.\textsuperscript{417} I would claim that von Wright’s main task in the philosophy of mind can be described as an attempt to explore the intelligibility of a reductive view. The measuring stick against which the intelligibility is judged is the language of everyday.

Although von Wright provides his own unorthodox theory of the mind, it seems to me that this, in his opinion, is not really the task of the philosopher. The task of the philosopher is

\textsuperscript{414} See Davidson, 1987 and 1990a.
\textsuperscript{415} At the same time it should be acknowledged that the \textit{a priori} status of the cause-law thesis is obscure. If the thesis is empirical then the main evidence for Davidson’s monism is empirical.
\textsuperscript{416} On the question of whether psycho-physical identity is an intelligible idea, von Wright’s views come close to the recent views of Putnam (1999) who has become highly critical of modern identity theories despite defending a version of the identity theory earlier.
\textsuperscript{417} A similar kind of view is expressed, for example, by Hacker, 2001.
to evaluate the intelligibility of empirical mind–body theories. As a lucky consequence, a new view may occur. When discussing the question of whether a philosopher should be fully acquainted with the previous philosophical landscape when he addresses a philosophical problem, von Wright notes: “[…] it has seemed to me that a new solution to an already solved problem can be of value both because it may have been reached through a new type of approach and because it may direct future research towards a new direction.”

In the preface of *In the Shadow of Descartes*, von Wright admits that in the philosophy of mind he often “had to go his own lonely way” because the already explored positions were something with which he simply could not agree. In hindsight, it can be seen that von Wright arrived at conclusions which have been suggested also by other philosophers of mind. It could be claimed that many philosophers of mind are exploring the intelligibility or unintelligibility of different positions. But there are essential differences in the background assumptions from which these philosophers proceed, and it seems to me that von Wright’s approach also challenges these assumptions in a productive way.

Through conceptual analysis, a philosopher attempts to determine whether the acceptance of physicalistic monism leads to incoherence, to a situation where some claims about the mind do not make clear sense. If this happens, the conclusion should be that the idea of monism is not intelligible because no sense can be made of what it means for such a position to be true. Even though von Wright’s main project in the philosophy of mind seems to be, in a somewhat Wittgensteinian spirit, the exploration of the intelligibility of different positions, he also wants to provide a general view which is not based on idle neuroscientific speculation but respects scientific facts. He argues that on such a scientific foundation a philosophical view which is fresh and healthy should be built. Such a view is free from metaphysical speculations. Needless to say, from a more empirically oriented perspective von Wright’s own views would most likely amount to hopeless speculations without much value for the modern philosophy of mind. Indeed, sometimes von Wright seems to offer speculations which do not fit well with his Wittgensteinian view according to which philosophy should not be speculative. Von Wright says things like: “The neurophysiological details… are… for all I know, largely obscure even to science. I have no idea what they are. Nevertheless I shall give free reign to my imagination in trying to think

---

418 Von Wright, 1999, 18.
what they may be.”

Or, with respect to these neurophysiological details: “I do not say that it is so – only that I fail to see why it could not be so”, and “Again, I do not say that it is so – only that it may be so.” It is unclear to me what the purpose of these speculations is supposed to be; they cannot seriously compete with empirical views. Of course, on von Wright’s understanding about the purpose of philosophy, philosophical views should not attempt to compete with empirical theories. But how do these speculations relate to empirical mind–body views? They certainly appear frustrating or irrelevant to philosophers who approach these problems in a more empirical vein. Those philosophers who insist that one should get out of “the narrow canyons of the commonsense conception of the world”, and that in philosophy there has been too much thought and too little experiment, do not value von Wright’s work due to its speculative and conceptual character. Likewise, von Wright claims that the modern trend in the philosophy of mind based on the findings of neuroscience is metaphysical, scientistic, and violently speculative. It is empirical metaphysics without real grounding except the scientific facts, which give rise to the very philosophical speculations it deplores. The facts do not justify all the philosophical conclusions which are drawn from them.

If a form of materialist monism is accepted, then its “postulational nature” and its status as a “scientific research program or heuristic idea” should also be recognized. Von Wright claims that there is and cannot be any proof for materialism; its truth or falsity remains beyond empirical evidence. Materialism is immune to criticism only in virtue of its being accepted as a postulate concerning the constitution of the world and concerning what should count as a scientific view of things. This is certainly not the way that Davidson would see the issue. It is perhaps somewhat surprising that von Wright seems to be closer to the type–type identity theorists than Davidson, although only in the sense that for him the truth of monism is an empirical issue. Von Wright agrees with identity theorists such as Smart and Place in their view that the identity theory is, above else, a scientific theory, or better still, a hypothesis. I have shown how Place himself contrasted this view with that of Davidson. But von Wright does not subscribe to materialism, as the type–type theorists do, just because it is the most plausible view from the scientific perspective. Kaila’s influence is thus not the only reason to defend monism. Von Wright mentions also the conceptual convenience which comes from the

---

420 Von Wright, 1989b, 28.
421 Ibid.
422 Churchland, 1988, 3.
423 Von Wright, 1998, 150.
424 Von Wright, 1989b, 29.
identification of the mental with the physical. He notes that by accepting materialism one does not need to “tamper with the principles of physical science, i.e. with what I shall call the Causal closedness of the physical order of things.” But, as I see it, this conceptual convenience is something which can be gained only after the principles of physical science have been accepted. In this sense, it could be claimed that an important part of the “conceptual convenience” is particularly the avoidance of conflict with a scientific world view. Should the resulting convenience actually be understood as showing the acceptance of a naturalistic view? For example, the problem of mental causation is a problem against the background of physicalism, and an attempt to fit mental causation conveniently into this context already assumes the truth of physicalism. The problem becomes real only if the truth of physicalism is assumed. One benefit of robust physicalism is that the problem of mental causation vanishes. There cannot be a problem of mental causation in the framework of physicalism. Should this count as a convenience? Kim, for example, has noted that the problem is not whether mental causation is true. The problem is how mental causation is possible.

Von Wright seems to accept the principle of the causal closure of the physical, although he does not explain why such a view should be accepted. A hint can, however, be found: von Wright talks about the assumed causal closedness of the physical world-order. This is a wise choice of words; there certainly is no proof that causal closedness is true. The principle of causal closure has the same postulational status as materialism. The assumptions are closely connected; it could be argued that the assumption in favor of materialism follows from the assumption that closure is true. One commentator, Friedrich Kambartel, has suggested that on von Wright’s view the Principle of Causality is a norm for scientific research and not an empirically confirmed statement about the structure of the world as it really is. This resembles Kim’s observation that it is not an empirical issue that physicalism involves the idea of the causal closure. I believe we can agree with Kambartel’s interpretation given von Wright’s view:

That every phenomenon has a cause, or that nothing can happen if not caused by something else is not an empirical truth to be established by observation or experience. Nor is it a necessary (conceptual or logical) truth established by rational argumentation… search for the causes of things has turned out immensely profitable.

425 Von Wright, UP, 1. The text has appeared in German. In the bibliography the German version is referred to as von Wright 2001a.
426 Von Wright, 2000.
427 Kambartel, 2005.
428 Von Wright, UPd 7.
This clearly resembles Davidson’s view, although Davidson would perhaps claim that the truth of the nomological character of causality can be established by rational argumentation. But, as I have shown, the consequent conceptual truth which Davidson arrives at concerns the relationship between the concept of cause and the concept of law. Whether something can actually happen in nature without a cause is a question to which Davidson’s argumentation gives no answer. Is an event without a physical cause an impossibility on a Davidsonian account? I shall not take a stance on this, but a question worth considering is: what actually speaks against events without causes?

What is the step from the principle of causality, understood as a norm, to causal closure? Why could there not be non-physical causes or non-physical effects? Von Wright’s simple answer is that their existence would conflict with our current scientific understanding concerning the structure of reality. The principle of causal closure has been used, often in a straightforward manner, in the argument for physicalistic monism. Von Wright writes, for example: “Sensations… have a causal origin in some physical event.”

Often this of course seems undeniably to be the case and a reasonable minimal physicalist would claim that this is always the case. But, if an argument for physicalistic monism is lacking, then the general claim is unwarranted, at least with respect to those sensations which, like hallucinations, seem not to have a clear physical cause. The view that the origin of sensations is a physical event is a view well confirmed by common-sense and science. But the philosophical question of whether sensations must have such an origin remains unresolved.

People often have sensations which seem not to, and sometimes do not, have any identifiable cause outside their bodies. These sensations do not have a cause which could be described as physical without begging the question. A sensation without an extra-bodily cause is perhaps the best example of a non-physical event; such an event does not fulfill the criteria which have been set for physical phenomena. If it is accepted that sensations themselves are physical phenomena, and if the causal closure of the physical is also accepted, then the assertion of the causal origin of all sensations can be given as a conclusion. But this is not an argument for monism, since the physical nature of sensations is already assumed. A different strategy is to follow the easy route of AM. If mental–physical interaction and closure

---

429 Von Wright, 1998, 55.
430 The physicalistic claim is that the causal origin of those sensations which do not have a cause outside the body lies inside the brain. This one can accept if one accepts the view of modern science. Yet, sensations have not been located in the brain, and the claim that they must have a physical cause is not based on the facts that are known about the brain. If it is based on anything it is based on the acceptance of the kind of postulate that von Wright suggests. These sensations, which do not have an extra-bodily source, may well have their origin in the brain, but this is question-begging if the question about the truth of physicalism is at stake.
are accepted, monism follows. If it is accepted that sensations have a physical cause and whatever has a physical cause must itself be physical, then monism follows.

Although von Wright does not have a special interest in giving an argument for physicalistic monism (and nor is he especially interested in defending such a position), he notes that there is a way to argue for its \textit{a priori} nature. This is done by sticking to the \textit{postulate} concerning the “constitution of the world” – in this case concerning the relationship of mind and matter. If this is done, the postulate becomes “(like a) logical necessity or an \textit{a priori} feature of reality.”\footnote{Von Wright, 1998, 149.} In my opinion it is not entirely clear how the expressions “logical necessity” or “\textit{a priori}” should be understood here, but von Wright’s suggestion that the postulate should be thought of as “a demand of reason” that is not an empirical truth established by observation or experience nor a necessary (conceptual or logical) truth gives the best suggestion for how the \textit{a priori} nature of psycho-physical identity should be understood.\footnote{Compare to Davidson’s (1995, 202) view about the cause-law thesis: “[…] the thesis clearly is not a pronouncement of ordinary logic, nor can it be established empirically.”} The “demand of reason” is something the nature of which, as far as I can tell, cannot be clearly defined. It seems to me that, for example, in the case of the mind–body problem there could certainly be conflicting “demands of reason”, and it is not easy to say how a conflict between them could be settled. Von Wright’s suggestion is nevertheless interesting because according to it, psycho-physical identity cannot be a logical necessity; yet it is not an empirical truth either. In a sense, this leaves the truth of physicalistic monism hanging in the air. If the question about the truth of monism is an empirical question, the problem loses its appeal as a philosophical problem. Perhaps this would demonstrate that a problem which once was philosophical has been converted into an empirical problem. But empirical research has not shown decisively whether or not mental phenomena are physical, and there are issues related to the mind–body problem which seem to resist, in principle, an empirical solution – like the question of whether humans have immortal souls which survive the death of their bodies or whether freedom of the will is possible. The question about the existence of soul is not a scientific question, which is not to say that it could not \textit{become} a scientific problem. The question about the possibility of immortal souls is (perhaps) a philosophical question and a dismissal of its relevance on scientific grounds shows a narrow understanding of the problematic human condition. Deep problems which are not scientific problems are nevertheless \textit{real} problems.

\textit{Postnote:}\footnote{Compare to Davidson’s (1995, 202) view about the cause-law thesis: “[…] the thesis clearly is not a pronouncement of ordinary logic, nor can it be established empirically.”}
What has been said has provided a grasp of the nature of physicalism that von Wright can be interpreted as supporting. He accepts a form of monism because it is in harmony with the current state of scientific knowledge. He does not give other arguments as to why monism should be accepted. Von Wright does not provide the kind of argument structure that Davidson does. He merely states how things are, or have to be, if a scientific world view is to be accepted and further suggests – without much argument, but following Kaila’s lead – that such a view should be accepted. I believe that the details of von Wright’s position, however, need clarification. He rejects interactionism and thereby agrees with “various forms of materialism”.\footnote{Von Wright, 1998, 149.} But it has become clear that materialism comes in many forms. Von Wright thinks, for example, that there is a necessary simultaneity between mental and physical phenomena and thereby calls his position “a kind of materialism”.\footnote{Von Wright, UPd, 7.} As I will show in the next section, the necessary simultaneity between the mental and the physical is best understood in terms of supervenience. The reasons for arguing for such simultaneity are the ones mentioned already; conceptual convenience and the existing empirical evidence.\footnote{Given what has been said before about the postulational nature of materialism, it is not entirely clear what necessary simultaneity can mean on von Wright’s account.} For these reasons, von Wright would accept the claim of “no mental difference without a physical difference”, and I believe that how this claim should be understood can be explicated by invoking the concept of supervenience.

We can conclude that von Wright shows an ontological bias towards materialism, just as Davidson does. He has sympathy for a view that the world is made of one “stuff.” But then the relevant question becomes: What form of materialism does von Wright propose? What is the nature of this “stuff”, and what follows for the philosophical questions about the mind if the world is made of it? Some commentators have noted von Wright’s clear bias towards materialism. De Caro writes: “From an ontological point of view von Wright is a monist (a physicalist or materialist monist)…”\footnote{De Caro, 1999, 125.} and lists the physicalistic assumptions made by von Wright, namely: “[…] ontological monism; the idea that there is some kind of neurological correspondent of every mental event; the thesis that only physical properties can be involved in causation.”\footnote{De Caro, 1999, 127.} These, incidentally, are exactly the assumptions made also by Davidson. Emiliani agrees with de Caro when he notes that von Wright’s view the world of the mind is the same as the world of the body,\footnote{Emiliani, 2005, 152 and Emiliani, 2001.} yet he claims that von Wright’s overall position is best
described as “neutral stuff monism”. Both de Caro and Emiliani note that von Wright succumbs to conceptual dualism or to dualism in relation to properties. This shows that the way in which von Wright’s position is interpreted is essentially the same as Davidson’s; physicalistic or neutral monism combined with a dualism of concepts or properties. It is curious that whereas Davidson’s reasons for monism have been under live discussion, von Wright’s reasons for monism are not considered by the few commentators who have studied his work. Davidson has been variously charged with the claim that his position is overtly physicalistic or not physicalistic enough. In von Wright’s case, no such charges have been presented. One reason for this is, of course, that von Wright was not a real participant in the discussion which is active in the modern philosophy of mind; his views were treated with silence, and perhaps they were considered not worth addressing. What I find surprising is that those who are interested in von Wright’s philosophy have not considered his reasons for monism. I suspect that these commentators uncritically accept a form of physicalistic monism, which needs no further argument than the remark that all other alternatives would amount to “supernaturalism”.

2.2.1 Von Wright’s supervenience

Von Wright defends the view that reality consists of one kind of stuff, and yet he argues for the irreducibility of the mental. The acceptance of mental–physical supervenience suggests that a form of token-physicalism comes closest to the best description of von Wright’s position. Against the interpretations of Emiliani and, for example de Caro, I argue that von Wright accepts supervenience and that it has an important role in his argumentation. There is actually clear textual evidence for this conclusion, but it is ignored by the commentators. De Caro is among those who claim that von Wright rejects supervenience. I agree that this is also a sensible interpretation. Von Wright for example explicitly says: “I... reject the quasi-causalist idea of the mental as something ‘supervenient’ on the material.” My claim is, however, that if von Wright really rejected supervenience, the nature of his physicalism would be a mystery, since supervenience is often interpreted capturing the essence of minimal physicalism. Rejecting supervenience would mean that there could be a mental difference without a physical difference, and what would be left of physicalism if this were true? Being a pedantic reader of von Wright, I want to note that in a reprint of the above article (in which

---

440 Von Wright, 1994a, 109.
the idea of supervenience is denied), that very passage is removed, perhaps suggesting that von Wright changed his mind.\textsuperscript{441} Of course this in itself does not show that von Wright came to accept the idea of supervenience. However, von Wright’s last statement on supervenience is:

\[\ldots\] one can replace talk of body–mind and mind–body interaction by talk of causal connections between physical events only. Using terminology now in fashion, one can say that... causal relations between mental and the physical phenomena are \textit{supervenient in time} on causal relations the terms of which are physical phenomena only. The suggested position comes as “near” the psycho-physical identity theory as is, in my opinion, “logically permissible.”\textsuperscript{442}

Von Wright also notes that the causal closedness of the physical world-order (nature) can be safeguarded by accepting the following: “Whenever a physical event P can be correctly called cause (effect) of a mental event M, there exists another physical event F such that the duration of M is included in the duration of F, and such that F is the effect (cause) of P.”\textsuperscript{443} In essential respects, this is how reductive materialism would formulate supervenience and how this position would treat the problem of mental causation.

But how should we understand these claims given that von Wright seems also to reject the idea of supervenience? A simple explanation would be that von Wright changed his mind. This could be a possible explanation, given von Wright’s own admission that with respect to certain questions his views went through significant revisions. I claim, however, that by rejecting the \textit{quasi-causalist} idea of supervenience von Wright criticizes those, targeted also by Kim, who claim that the causal efficacy of the mental could be simply and non-reductively explained by invoking the concept of supervenience. Davidson, for example, seems to be among these philosophers.\textsuperscript{444} Kim has suggested that perhaps \textit{supervenient causation} would be something which could explain the mystery of mental causation in a non-reductive way.\textsuperscript{445} If supervenience explains the mind–body relation non-reductively, why not invoke supervenient causation to explain non-reductively the relation of mental causation? Why not say that a mental event M is a \textit{derivative cause} of a physical event P in virtue of its supervenience on P, which undeniably \textit{is} the cause of P? This would be an easy solution, but would it grant “real causal powers” to M? Kim has become critical towards his earlier view and has later claimed that this kind of “trick” would not secure the causal status of M. He now

\textsuperscript{441} See Von Wright, 1994, 148.
\textsuperscript{442} Von Wright, 1999, 30.
\textsuperscript{443} Von Wright, UPd, 6.
\textsuperscript{444} It is at least possible to interpret him in this way. See Davidson, 1993d.
\textsuperscript{445} Kim, 1984.
claims that supervenient causation could as well be called “pretend” or “faux” causation, and that a model in which the causal powers of M are meant to be secured on the basis of its supervenience on P₁ “could… be philosophically pernicious if it should mislead us into thinking that we have thereby conferred on M, the mental event, some real causal role”.\textsuperscript{446} I think von Wright’s rejection of quasi-causalist supervenience is motivated precisely by his belief that its acceptance would lead to philosophically pernicious results. He argues against the idea of supervenient causation on the same grounds as Kim, who is a robust physicalist. Kim now claims that supervenience is best understood as a reductive relation, and von Wright – although highly critical towards reductive positions in the philosophy of mind – seems to come very close to this view.

Although quasi-causalist supervenience should be rejected, von Wright claims that supervenience nevertheless describes the most “permissible” psycho-physical relation. The expression “supervenient in time” is not used in the discussions of supervenience, and therefore it would be a matter of speculation to say how von Wright’s formulation of supervenience should be exactly understood. Besides von Wright I have not come across other philosophers who have used the expression “supervenient \textit{in} time”. Von Wright’s views about supervenience remain brief and vague. He formulates supervenience in terms of temporal and not spatial coincidence, and does not discuss concepts like “realization”, “determination” or “constitution” that are usually associated with supervenience. In von Wright’s case, the notion does not explain much. The same can actually be said of Davidson’s use of the concept. The best description that can be found from the writings of von Wright about the relation of mental and physical event is the claim that it is a part-whole relation; the mental is (concretely) part of the physical, but the relation cannot easily be understood as a relation of causality or a relation of \textit{identity}. What von Wright above all seems to emphasize by his use of the concept of supervenience is that “mental things” are \textit{not extended in space}, \textit{yet they last over a period of time}. He notes that it is common to say that mental things may be located in time but have no extension in space.\textsuperscript{447} This being said, von Wright’s supervenience sometimes has a more substantial, ontological, aspect when he writes, for example, that a volitional act is \textit{included} in cerebral processes (in my view this suggests a \textit{spatial} simultaneity) – or when he defines a reason as a mental episode which is a temporal \textit{segment} of a neural cause of a bodily movement.\textsuperscript{448} Whatever claims about supervenience von

\textsuperscript{446} Kim, 2003, 148.
\textsuperscript{447} See for example Von Wright, 1997.
\textsuperscript{448} Von Wright, UPd.
Wright makes, it is clear that his views on are not based on empirical evidence. The claim about the part–whole relation is not a result of brain-research; von Wright’s view is based on conceptual investigations. Likewise, the claim that a mental thing is not extended in space would sound obscure to philosophers inspired by cognitive neuroscience. But here again, von Wright is making a conceptual point by claiming that the concept of a mental state is such that spatial location is not part of it. The exact location of a mental state does not make sense if “mental state” is understood in a certain way – that is, in terms of the language of the everyday. Ontologically speaking, a mental thing’s extension in space is nevertheless something that von Wright, because of his rejection of dualism, cannot in my view deny.

Von Wright’s views about supervenience have been largely neglected by commentators. De Caro speculates that von Wright could accept “Davidson’s epiphenomenalism” if it weren’t for Davidson’s acceptance of the supervenience thesis. Emiliani claims that according to the usual understanding of supervenience: “Descriptions in mental terms… are held to supervene on physical ones…”, but that in the case of von Wright: “an ascription of a mental state is not a description of states and processes of [a neural] kind” and that “token-identity is rejected – not even an individual action is identical to a set of physical processes…. As I have noted, it is difficult to say how the claim that causal relations between the mental and the physical are supervenient in time on the relations the terms of which are physical phenomena only should be understood. But it is not clear, pace Emiliani, that this way of stating the situation differs in any important respect from the view that “descriptions in mental terms supervene on physical ones”.

Von Wright’s emphasis is on the temporal identity of a mental event M and a physical event P. “Supervenience in time” means that events temporally coincide or that event M’s temporal duration is “included” in that of P. These expressions are rather obscure – how to understand the inclusion of temporal duration? I think this view does not imply that M would have to be spatio-temporally identical with P; that is, it does not imply that “supervenience in time” is compatible with the view that M is mysteriously spatially distinct from P. It is possible that M and P do not occupy exactly the same region of space, or more dramatically, that M does not occur in space at all. This would be a version of Cartesian dualism. Von Wright notes that: “The nervous and the sensational (physical and mental) coincide without

---

449 De Caro, 1999, 128, fn. 9.  
451 Von Wright even claims that the duration of mental states is necessarily included in the duration of the corresponding physical states and events. See, von Wright, UPd.
being identical.” But what this “coincidence without identity” precisely means is unclear, because von Wright also writes that if there exist a “chain of reasons” and a “chain of causes” which converge in an action, then “[F]rom the point of view of their ‘substrate’, i.e. their robust, spatio-temporal reality, there is only one ‘chain’.” I think that the expression “robust, spatio-temporal reality” sits uneasily together with von Wright’s conceptual dualism, and shows a certain kind of bias towards the physical. If there is only one chain, the result is that action is nothing over and above its physical aspect, if by “over and above” we understand some event or entity in the physical world. Doesn’t this imply the identity, and not a mere coincidence, of the chains because there is only one chain? It is obvious that the “one and only chain” of which von Wright speaks is physical, since he claims that the idea of mind–body interaction can be replaced with the idea that causal connections occur only between physical events. This sounds similar to the position of physicalism that is usually attributed to Davidson. It can be concluded that it is plausible that M’s supervenience in time on P should be understood as implying a spatial supervenience as well, although this way of formulating the idea of supervenience is perhaps also compatible with the view that M and F are spatially distinct “entities”. Supervenience in time leaves room for the possibility that the “base” entity and the supervening entity are spatially distinct, although if a change in the base entity is required for the change in the supervening entity, then there has to be (if we exclude the possibility of pre-established harmony and mere co-incidence) some kind of causal interaction (or identity) between the entities. Could it be claimed that all events which occur at the same time are “supervenient in time” with respect to each other? No, at least not according to the usual understanding of supervenience, which sees supervenience as a relation of dependence or determination. This being the case, the supervening entities must be in some kind of hierarchical relation to one another. On the other hand, all the talk about “base entities” and “supervening entities” sounds obscure if supervenience is thought to concern predicates and not properties. If the predicate of goodness supervenes on natural properties, is “goodness” an entity distinct from the properties on the basis of which this predicate is applied? These problems will be discussed in section 2.3.

How should we understand the claim that an ascription of a mental state is not a description or ascription of a brain-process? This claim, put forward by Emiliani, is of course...

---

454 As one commentator (Kivinen, 1998) has noted, the “exorcism of Descartes” is not so easy because Descartes could counter: The action may not be anything over and above its physical aspect in the physical world, but how about the mental world?
difficult to understand if one accepts minimal physicalism and accepts that, whether or not it can be shown, “deep down” mental states are brain-states. It is equally clear that whether they are is one of the important questions under consideration, and the search for a philosophical answer to this question continues. Von Wright’s agreement with various forms of materialism is a good, perhaps compelling, reason to think that he would take mental states to be states of the brain. Given this, the claim that von Wright rejects the thesis of token-identity is problematic. From the fact that a particular mental phenomenon cannot be identified with a physical phenomenon it does not follow that the former, when it occurs, is not identical with the latter. But sometimes von Wright draws exactly the conclusion that M cannot be identical with P because no description of P would be sufficient to identify M conclusively. Yet, von Wright suggests that it is “plausible”, although a “sheer hypothesis”, that there is a difference at the neural level – for example, between cases when a person feels pain and when he does not. The feeling of pain is a mental phenomenon; it is supposed to have a counterpart at the neural level, and as the standard formulation of supervenience claims, any change in the feeling requires a change at the neural level. I believe there is no reason to think that von Wright would deny the correctness of this formulation. A naturalist would, of course, argue that the claim that a neural difference is required for a mental difference is more than a “sheer hypothesis”: there has to be a difference. Von Wright’s reply is that this claim is a philosopher’s dogma, one dogma of materialism.

The talk of the neural “counterparts” of mental phenomena causes confusions as long as we stay in the context of an all-encompassing materialism. If O-physicalism is accepted there does not – ontologically speaking – occur any mental phenomenon M which would be something “over and above” a neural phenomenon N. There are no distinct entities to be correlated. But I think it is possible to understand what all the talk about “counterparts” means, because in the course of everyday life, beliefs and other mental phenomena are attributed to others without any knowledge about their brains. There seem to be two distinct phenomena, mental and physical, and mental phenomena exist, so to speak, in their own right. Their relation to the physical world can be considered. It is a plausible view, according to naturalistic philosophy at least, that the phenomena attributed take place in the brain. Von Wright notes:

---

455 This seems to be Emiliani’s reason for claiming that von Wright rejects token-physicalism.
Sensations and perceptions, recollections and thoughts, reasons for action may all have distinct physical “counterparts” in the neural system without which those psychological phenomena would not occur [in] the person who has them.\textsuperscript{457}

Here von Wright notes merely that mental phenomena \textit{may} have distinct physical counterparts.\textsuperscript{458} This modal claim, which can be formulated as “it is possible that…” should be understood in the context of von Wright’s view, according to which the logical necessity of mind–body identity cannot be established. By emphasizing that it \textit{may} be so, von Wright is asserting his “negative point” according to which it does not \textit{have} to be the case that the identity-claim is true. But, or so I think, von Wright is also hinting at a more “positive point”, according to which it is possible that mental phenomena may turn out to be distinct, \textit{clear} or \textit{specific} physical phenomena. Isn’t this precisely the view that mental is something supervenient on the material, and does it not assert the claim of token-physicalism? Yes. Something in the brain is required for the occurrence of psychological phenomena; no mental difference without a physical difference. This being the case, it is no wonder that von Wright writes: “I would like to say that in feeling pain we immediately experience (sense) something that goes on in our nervous system”,\textsuperscript{459} and that “Could we… not say that pain as physical phenomenon is the nervous processes which (or part of which) we experience as pain?\textsuperscript{460} This certainly gives the impression that, pace the non-physicalistic reading of Emiliani, an ascription of a mental state is \textit{exactly} a description of states and processes of a neural kind.

Von Wright’s criticism of supervenience does not show that he would reject the idea of mind–matter supervenience, and the textual evidence shows that it actually has an important role in his argumentation. Because many modern philosophers want to avoid dualism, they may accept mind–matter supervenience while claiming that such a relation does not make good sense since it is not clear what the concept says about the relation between mind and body. This “solution” is expressed in Hilary Putnam’s views when he claims that

\textsuperscript{457} Von Wright, 1989b, 29.
\textsuperscript{458} As noted already, “distinct physical counterpart” is an obscure statement if the position is meant to be a form of physicalism. It should be noted, though, that the obscurity of the claim depends on the meaning of the term “distinct”. Does it mean something like “clear” in this context, or is it meant to mean something like “separate”? A clear physical counterpart of a mental phenomenon does not sound as obscure as a distinct physical counterpart.
\textsuperscript{459} Von Wright, 1998, 165. Consider also: “When we hear, see, feel, taste or smell, we \textit{sense} the nervous processes, or a part of them which are \textit{caused} by the sensory stimulus… in a sense, the primary objects of sensations are nervous processes.” (Von Wright, 1997a, 152.) This being said, it must be stressed that von Wright apparently went through a change of mind, since in von Wright, 2000 he claims that what we hear is the \textit{source} of the sound in the outer world, not the nervous processes which are produced by the sound. As far as I can tell one can advance either one of these theses, that we hear the nervous processes or their source. Moreover, in von Wright 2000a the sensation under discussion is exclusively \textit{hearing}. It is therefore not clear whether the views would apply to all sensations, not to mention “higher” mental phenomena like beliefs.
\textsuperscript{460} Von Wright, 1998, 167.
“mental events aren’t ‘identical’ or ‘not identical’ with physical events”.\textsuperscript{461} This is an obscure statement, but the rationale behind it seems to be that the notion of identity has not been given a clear sense in the mind–body context, and that therefore the apparent identity between the mental and the physical is “a criterionless and \textit{sui generis} sort of “identity””.\textsuperscript{462} This resembles Malcolm’s Wittgensteinian view when he notes that the claim according to which mental phenomena have neural counterparts is “without sense until it is given a sense – and giving it a sense will not be easy”.\textsuperscript{463} It is, so to speak, up to us what kind of sense will be given to this view and what kind of status will be granted to it in our conceptual scheme. The sense that \textit{can} be given to the idea that mental phenomena have physical counterparts is constrained by certain facts; whether the constraining facts are taken to be essentially scientific facts or the facts of everyday life is based on a choice which is not easy to defend.

The position of Putnam and Malcolm captures part of von Wright’s view, since according to him the assertion that mental states are \textit{identical} with their physical counterparts is “a serious confusion”. In some sense, the mental and the physical are parts of the same reality; mental and physical ascriptions capture the same part of reality, the same spatio-temporal region or, as von Wright would say, at least a part of the same temporal region. Von Wright asserts that the view according to which mental states and events are identical with their counterparts in the neural system has \textit{some} truth in it. But the obscurities of token-identity make it problematic to say that M and P are identical. Non-reductive physicalism faces serious problems in its claim that mental phenomena are token-identical to physical phenomena. I will return to this problem in section 2.4, where the details of von Wright’s and Davidson’s views about token-identity are discussed.

As far as von Wright’s physicalism is concerned, there is enough textual evidence to warrant the conclusion that he would accept the idea of supervenience, and thus a form of minimal physicalism. This may not come as a surprise to commentators, although the general view seems to be that, because of his conceptual dualism, von Wright should not be described as a physicalist at all. What may be surprising is that von Wright’s acceptance of monism has in fact robustly physicalistic, or even eliminativist, consequences, the most evident of them being the denial of mental–physical or physical–mental causation. I would thus argue that,

\begin{flushright}
\textsuperscript{461} Putnam, 1999, 75.
\textsuperscript{462} Putnam, 1999, 37. Putnam criticizes especially Davidson’s criterion, which he takes to be circular. How the intuitions of philosophers differ is shown well, for example, by the fact that Fodor complains of not understanding Putnam because the former has “no clue what it is to give a sense to a notion; the notion of giving a sense to a notion hasn’t been given a sense….” (Fodor, 2000). Fodor does not problematize the question of whether a satisfactory identity criterion can be given for mental phenomena. According to him, the question \textit{where} in the brain mental phenomena occur is not interesting.
\textsuperscript{463} Malcolm, 1989, 36.
\end{flushright}
pace Emiliani, there are reasons to think that von Wright’s conclusions are in fact more physicalistic than Davidson’s. Indeed, in an unpublished manuscript von Wright himself speculates that his position, by eliminating the mental, has some likeness to the position of eliminative materialism that he associates with Patricia Churchland. This is a very interesting claim, which I nevertheless find extremely puzzling. The claim is especially puzzling given that von Wright identifies eliminative materialism as one of the most vulgar examples of the scientistic tendency in the modern philosophy of mind. What von Wright’s view shows, however, is that the interpretation according to which he stands firmly against robust physicalism is at least partly incorrect; von Wright cannot be thoroughly non-reductivist if he sees a similarity between his own views and those of Churchland. One way to arrive at an epiphenomenalist conclusion, which claims that the mental aspect of human life is causally inefficacious, is through defending the view that only physical properties can participate in causal relations. This, the causal primacy of physical properties, is one of the features of von Wright’s physicalism. My claim is that there are reasons to think that the strength of von Wright’s physicalism is actually quite high. He says things like: “It would be surprising if... variations of our experience did not answer to determinate variations in some neural processes.” Or: “Could not all... different ‘states of mind’ be reflected in characteristically different neural states? The idea seems (to me) very natural, empirically plausible – maybe even logically compelling.”

The first quote raises an interesting question concerning von Wright’s physicalism. If variations of experience answer to determinate variations in neural processes, is the position actually closer to type-physicalism than token-physicalism? Does the view mean that a mental state M is always, or at least with high frequency, correlated with a specific (“distinct”) neural state N? This would be a stronger claim than Davidson’s, who argues explicitly against psycho-physical laws and therefore against the view that there are determinate variations at the neural level corresponding to the variations of experience. Given von Wright’s other views on the subject it is unclear how the expression “determinate variations” should be taken, but the formulation definitely raises questions. Whether generalities at the level of correlations between mental and physical phenomena can be found seems, in the end, to be an

---

464 Von Wright, UPa. The reference is made in a sketch which eventually came to be the paper “On Mind and Matter” (von Wright, 1994).
466 This is the standard accusation of epiphenomenalism laid against Davidson. I will consider its plausibility in chapter four, where I discuss the question of why, according to my interpretation, von Wright’s views are closely related to epiphenomenalism.
467 Von Wright, 1998, 117.
468 Von Wright, 1998, 133.
open question according to von Wright. He often emphasizes that only empirical research can provide an answer to this question.

The second quote is also interesting. A view that states of mind may be reflected in characteristically\textsuperscript{469} different ways may be empirically plausible, but what does it mean to say that this view is “very natural”? It is certainly not a truth of everyday language, which is the transcendental bedrock that, according to von Wright, must be respected. The same question can be raised with respect to expression “logically compelling”. In what sense can the idea be logically compelling? At times von Wright seems to claim just the opposite: monism is contingently true. It is not a contradiction in terms to think that a contingent position is logically compelling, but it is also unclear what the expression can be taken to mean in this context. Perhaps in von Wright’s view the claim that mental phenomena are reflected in characteristically different neural states is something that it is rational to accept. This, it seems to me, would nevertheless undermine some other claims of von Wright – especially his emphasis on the “primacy of the obvious”.

Von Wright accepts the token identity thesis to the extent that, because of physicalistic monism, it is either trivially true or, because of conceptual dualism, it is nonsense. The claim that token-identity is trivially true seems to be an acceptance of global supervenience, whereas the claim that token-identity is “nonsense” resembles the views of Putnam and Malcolm. To say that the token-identity thesis is trivially true is to assert that whenever something mental occurs, something takes place in the brain. This von Wright accepts, because the alternative would amount to accepting a “magical”, non-scientific view of the mind–brain relationship. But an attempt to go beyond this, an attempt to identify mental states with physical states, would lead to nonsensical claims. I believe it must be concluded that in von Wright’s philosophy of mind, there is no room for a view according to which a mental phenomenon would not be token-identical with a physical phenomenon in the trivial sense. This is, of course, no wonder given von Wright’s materialism; token-identity captures the essence of non-reductive physicalism. From what has been said, I cannot but conclude that those who see an important difference between von Wright’s and Davidson’s versions of materialism are wrong. Their forms of materialism are equally strong or equally weak. Moreover, the view that von Wright rejects the idea of mental–physical supervenience does not capture the “spirit” of von Wright’s views on the relationship between mind and matter. Incidentally, Emiliani’s claim that “[…] von Wright’s conception basically departs from

\textsuperscript{469} Should this term be understood in the same sense as “determinate variations”? If so, this again suggests that the correspondences between the mental and the neural would be closer to type-identities than token-identities.
Davidson’s – as well as from several approaches to the mind–body problem focusing upon the notions of supervenience and emergence is suspect – not only with respect to the claim about supervenience, but also with respect to the claim about emergence. Von Wright specifically notes:

I am… inclined to speak of the sensation as neither caused nor causally efficacious but as being a ‘byproduct’ which emerges as a concomitant of the nervous process…On this view mental states and events are kind of… ‘emergent qualities’ of their ‘physical substrate’.

This formulation sounds paradoxically both very physicalistic and very anti-physicalistic. On the one hand, a view which denies the causal efficaciousness of sensations is a form of eliminativism, and such a position is usually interpreted as being the strictest form of physicalism. Von Wright’s reference to Churchland, who is openly eliminativist, cannot be ignored. On the other hand, a view which mentions a physical substrate and *emergent qualities* sounds like a form of emergentism, and is thus not a version of physicalism at all. If von Wright did subscribe to a form of genuine emergentism, this could also explain his non-standard use of the concept of supervenience. Supervenience was used by the British emergentist Samuel Alexander to describe a relation which holds between emergent properties and base properties. In 1920, Alexander writes:

The higher quality emerges from the lower level of existence and has its roots therein, but it emerges therefrom, and it does not belong to that lower level, but constitutes its possessor a new order of existent with its special laws of behaviour. The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I would prefer to say in less harsh terms, to be accepted with the “natural piety” of the investigator. It admits no explanation.

According to the emergentists, the emergent properties are genuinely distinct and additional to the base properties; there was a time when “supervenience” referred to the occurrence of emergent properties. Given von Wright’s use of the term “emergent qualities”, it could be thought that his understanding of supervenience would agree with Alexander’s formulation. However, we have seen that von Wright emphasizes the primacy of the physical in his

---

471 Von Wright, 1997a, 155.
472 Alexander, 1979, 46.
473 Whereas, for example Davidson, describes supervenience as a relation which guarantees that mental properties matter to causal relations, von Wright (UPb) describes supervenience as a relation which allows for the existence of supervenient properties, but only as *shadows* which neither have influence on the base properties, nor are influenced by them. How this understanding relates, for example, to Alexander’s understanding is an interesting further question, which I shall not consider here.
explanation of how the relation of supervenience should be understood. I believe that the interpretation according to which von Wright is very physicalistic instead of anti-physicalistic must therefore prevail. Yet, I am tempted to note that in Explanation and Understanding, von Wright likens emergence to a process of transmutation of quantity into quality. Through our understanding something physical in a new way, the physical takes on a quality which it did not have before. As a curiosity, it can be noted that in the Philosophical Investigations Wittgenstein wrote: “Our attitude to what is alive and to what is dead is not the same… if anyone says ‘That cannot simply come from the fact that a living thing moves about in such a such way and dead one not,’ then I want to intimate to him that this is a case of the transition from quantity to quality.” Should emergent qualities be understood as resulting from our capability to see aspects of the world in certain way, which in turn results from our attitude?

Von Wright’s own conclusion in fact is that perhaps his position could be best described as a form of emergentism or epiphenomenalism. These positions do not exclude a monistic physicalistic ontology but are perfectly compatible with it. In fact, as noted earlier, epiphenomenalism, a causal eliminativism with respect to mental phenomena, could be seen as a version of extremely strong physicalism. Von Wright notes that the necessary simultaneity of the mental and physical, and the consequent causal inefficacy of the mental, could be interpreted as a form of materialism which is nevertheless not reductive. But a non-reductive position can be a version of eliminativism. I have argued that von Wright’s reason for defending materialism is scientific intelligibility, and I have also shown that, in the case of the Churchlands and others, this position has led to the elimination of the mental. Whereas Davidson always insisted that his view does not lead to epiphenomenalism, von Wright admits that his position has this consequence. Interestingly, commentators who have accused Davidson’s position as endorsing a form of epiphenomenalism have largely ignored the relation between von Wright’s view and epiphenomenalism. Emiliani’s claim that von Wright’s position does not focus on the notion of emergence is incorrect, because it is this very concept that von Wright uses in describing the nature of sensations. No one has studied this side of von Wright’s philosophy of mind and the consequences to which it leads. I think this aspect could be developed further from two different directions. On the one hand, the claim that emergence is a process where quantity turns into quality could be further explored.

---

474 Von Wright, 1974, 135.
475 Wittgenstein, 1953, §284.
476 See, von Wright, 1997a, 1998 and 1999. Von Wright (1997a) notes that he is not acquainted with any positions of emergentism or epiphenomenalism that would be identical or even very similar to his position.
477 Von Wright, UPd, 7.
This would be a “dualistic” direction. On the other hand, it could be considered whether supervenience – understood as a reductive relation – deprives supervening entities of their causal powers and makes them emergent qualities which are something like byproducts of physical phenomena. This would be an “eliminativist” direction. It is fascinating that von Wright’s position allows for both readings. He did not develop these directions further, and for the purposes of this work it would not be beneficial to speculate about the possible directions. I hope to be able to develop these thoughts in the future.478

I think that three conclusions can be drawn. First, von Wright sees his position as a form of materialism. Second, he stipulates an important role for the relation, if not the concept, of supervenience, which comes as near to establishing psycho-physical identity as is possible. Third, he sees mental phenomena as emergent qualities of a physical substrate. All the essential claims of non-reductive physicalism are in place; monistic metaphysics (materialism), a critique of reductionism and an acceptance of conceptual dualism, a non-reductive relationship of mind and matter (supervenience) that the concept of emergence is meant to further clarify. This is in contrast to Davidson, who was careful not to use the concept of emergence as a description of the mind–matter relationship. Does my interpretation once more change the roles of von Wright and Davidson with respect to the strength of their physicalism? By accepting “emergent qualities”, von Wright would seem to be less of a physicalist than Davidson, who sees emergence as being “merely” a conceptual issue.479 Since von Wright does not elaborate on his use of the concept of emergence, a conclusive decision about the role it plays in his views about the mind–matter relationship cannot be drawn. The fact that he nevertheless uses this concept is interesting given the contemporary discussion about the (in)coherence of emergentism in the context of physicalism. The most severe criticism against emergentism from the perspective of physicalism has usually been that it fails to explain how mental causation is possible. I want to emphasize that von Wright is immune to this criticism, because he gladly accepts the view that, in a sense, mental causation can be reductively explained – the consequence being that mental phenomena turn out to be epiphenomenal by-products of neural phenomena.480 The combination of monism and “emergent qualities” is not an easy alliance, and the conclusion, epiphenomenalism, is of course dramatic.

478 Von Wright’s relation to epiphenomenalism is explored in Kuusela 2009.
479 For Davidson’s brief remarks about emergence, see Davidson, 1997a. For further consideration of the conceptual nature of emergence and a discussion of Davidson’s position see Pihlström, 2002.
480 The sense in which von Wright accepts this claim will be discussed in section 4.2.1.1.
By rejecting interactionism and accepting a form of materialist monism, von Wright solves the ontological problem created by Descartes. This is basically the same solution as Davidson’s if the plausible interpretation that AM is a form of physicalism is accepted. Insofar as both von Wright and Davidson are interpreted as being physicalists, there is no important difference between their views. Non-reductive physicalism captures the essence of their positions although, as I have shown, certain views of von Wright point towards a more robust form of physicalism. There is also room for a different interpretation according to which the positions of Davidson and von Wright are not in any clear sense physicalistic positions at all. The view that von Wright and Davidson could be described as defending a form of neutral monism will be presented in section 2.6. I think the fact that both philosophers can be interpreted as physicalists, non-reductive physicalists, and also as neutral monists is a very interesting similarity between them, since both von Wright and Davidson have arrived at their rather original conclusions in separate ways. The non-physicalistic suggestions of von Wright and Davidson provide an alternative which should be considered seriously when trying to come up with a “solution” to the mind–body problem. This is especially so if one is dissatisfied with the way in which robust physicalism deals with the question and is unwilling to accept the naturalistic view at face value. I believe that in the end it should be argued that the spirit of the positions of von Wright and Davidson goes against robust physicalism.

This being said, it cannot be denied that there are very good reasons to think that both Davidson and von Wright see physicalism as an answer to the mind–body problem. This is the conclusion of sections 2.1 and 2.2. I have argued that Davidson’s argument for physicalism is far from convincing and that von Wright does not actually give an argument for physicalism at all. Whereas Davidson has become known for his monistic theory of mind, von Wright merely states his allegiance to a scientific world view and its conviction that the nature of reality is physical. In this sense, von Wright’s views about monism are less interesting than Davidson’s, and his straightforward acceptance of monism is, in my opinion, surprising. At the same time, it is important to keep in mind that von Wright emphasizes the postulational nature of materialism and therefore strongly undermines the philosophical interest of the question of whether materialism is true. Perhaps one could conclude, as for example Stoutland does, that both Davidson and von Wright see minimal ontological physicalism as a position which just about any philosopher would accept, and that it therefore

---

481 As should be clear, the motivation behind Davidson’s philosophical argument is certainly, at least in part, the view that the results of science should be respected. But whereas von Wright is happy to rest with this attitude, Davidson’s argument for monism goes a step further and is this sense an important philosophical contribution to physicalism. Von Wright seems to think that it is not a philosopher’s task to make such a contribution.
does not need much argumentation for its defense. But as has become clear, the status of even minimal ontological physicalism is still a debated issue among philosophers, and a consideration of the non-physical alternatives does not make one automatically a “supernaturalist” – or if it does, it is time to consider whether opponents of supernaturalism are in fact restricting themselves to a narrow understanding of the term “natural”. If a generous enough attitude is taken, then nothing counts as supernatural. It is not clear that this would necessarily be an unacceptable position; perhaps certain forms of pragmatism could be seen as moving in this kind of direction. Conceptual clarification is what a philosopher is trying to achieve, and an open mind is required in order to evaluate the deep conceptual problems which still surround the mind–body problem in its ontological form. The plausibility of physicalism should be considered in an unbiased manner: it should be considered whether the claim that “all entities are physical” provides any substantial clarification of the question about the relationship between mind and matter – and if it does not, then it should be considered whether the claim is unsupported and should be rejected.

In the next two sections, I consider the problems of non-reductive physicalism. For the most part I will ignore other positions than Davidson’s and von Wright’s, because my purpose is to consider whether these two philosophers succeed in providing serious support for a physicalistic world view. In order to determine this, we must consider whether their position can solve two of the most serious problems which allegedly face non-reductive physicalistic views about the mind.

2.3 A problem for non-reductive physicalism: Mental properties

Philosophers who are minimal physicalists disagree on the question of whether there are non-physical properties. Non-reductive physicalists argue that mental properties are not physical properties; thus the label property dualism is often associated with non-reductive physicalism. In this section, I will consider how Davidson discusses the problem of irreducible mental properties. Von Wright’s views on the subject cannot be discussed in great detail because he has been silent on the issue. This being said, in an unpublished manuscript von Wright briefly discussed the problem of mental properties in a way which shows that he was thinking about the problem, although his views were never published. The actual result of the unpublished manuscript is von Wright’s “An Essay on Door-Knocking”, where the problem of mental

---

482 See von Wright, UPg.
properties is not given much attention. I will very briefly describe what von Wright did say about the problem, thereby showing what his position with respect to mental properties was. My purpose is to place the positions of Davidson and von Wright in the current context, the one in which the discussion about the status of mental properties is currently being carried on. I will do this by considering how well their views agree with the standard view about the nature of mental properties (in terms of which the contemporary problem is usually formulated).

The question about the status of mental properties has become an important part of the ontological mind–body problem for physicalists. The term “property dualism”, of which non-reductive physicalism is allegedly a form, is understood and used in different ways in the current philosophical literature. According to Searle, a property dualist endorses the following views:

1) There are two mutually exclusive metaphysical categories, of mental and physical phenomena, which constitute all of empirical reality.

2) Because mental states are not reducible to neurobiological states, they are something distinct from and over and above neurobiological states. The irreducibility of the mental to the physical is by itself sufficient proof of the distinctness of the mental.

3) Mental phenomena do not constitute separate objects or substances, but rather are features or properties of the composite entity, which is a human being.

This formulation illustrates well how the discussion of property dualism in the context of physicalism leads immediately to obscurities. What is a metaphysical category? What does it mean to say that a mental state is distinct from a neurobiological state? What is a mental feature or property of a human being? Davidson, for example, argues that mental phenomena do not constitute an ontological category. Von Wright agrees with this. If “metaphysical category” is a synonym for “ontological category”, then Davidson and von Wright do not defend the view (1), which Searle attributes to property dualists. It is thus incorrect to say, according to all property dualists, that mental properties constitute an ontological class of

---

483 Searle, 2002.
their own. If one endorses materialist monism, as von Wright and Davidson do, it is unclear what the distinctiveness of mental states from neurobiological states could mean. Searle correctly points out that the irreducibility of the mental is sufficient proof of the distinctness of the mental in some sense. But for many philosophers, the irreducibility of the mental is a conceptual question and therefore does not imply ontological or metaphysical distinctiveness. Finally, the claim that mental phenomena are features or properties of a human being is obscure. The term “feature” is sometimes used interchangeably with the term “property” but sometimes the expression “feature” is used to refer to something less substantial than a property. These terms are used unproblematically in the discussion about the nature of property dualism, but it is not clear what it means to say that being in pain is a feature of a person. In sum, we can conclude that in the context of materialism the meaning of “property dualism” is very problematic.

The task of coming up with a clarifying definition of property dualism has been taken up by many philosophers. Searle has described property dualism by describing the principles endorsed by a property dualist. According to Kim, non-reductive physicalism is committed to the idea that mental properties constitute “an autonomous domain” that resists reduction to the physical domain, and therefore mental properties should be understood as being distinct from physical properties. Mental properties are irreducible in the sense that they are not to be found among the properties of basic physics and are not reducible to such properties. What Kim actually means by “property dualism” is this: a property dualist would claim that higher-level features of the brain are not reducible to, or reductively identifiable with, the lower level properties. Curiously, this formulation makes Searle, for example, a property dualist, contrary to his own views. Biological naturalism would thus be a form of property dualism together, for example, with AM. This shows that in the contemporary philosophy of mind the boundaries of property dualism are not always clear. In my opinion this is not just a terminological question; the important question is what characterizes modern property dualism according to those who see the position as incoherent.

The question about the status of irreducible mental properties relates to the more general problem concerning the coherence of non-reductive physicalism. Kim has been one of the main critics of non-reductive physicalism, arguing that ontological physicalism and an

---

484 This is of course a matter of controversy. It is clear, however, that both von Wright and Davidson deny that such irreducibility would have consequences for a physicalistic ontology.

485 For Searle’s denial, see Searle, 2002. Searle (1994) notes that he has been described as a materialist and a dualist. The same has been said about Davidson.
autonomous mental domain make an incoherent combination. Kim writes that according to a non-reductive physicalist, it is possible to:

[...] assuage our physicalist qualms by embracing ‘ontological physicalism’, the claim that all that exists in spacetime is physical, but, at the same time accept ‘property dualism’, a dualism... insisting that psychological concepts or properties form an irreducible, autonomous domain.

Here Kim formulates the position of property dualism in a similar vein as Searle: a property dualist insists that psychological (mental) concepts or properties form an irreducible domain. At least here, Kim brings together concepts and properties. But surely the question of whether a property dualist is talking about concepts or properties makes a difference to his position. If this distinction is ignored, a serious confusion in the discussion about the nature of property dualism occurs.

Kim’s formulation nevertheless captures the spirit of Davidson’s position. Defined in the way Kim does, the label of non-reductive physicalism applies also to von Wright’s position. Psychological properties are not physical properties although everything is physical. Kim and many other contemporary naturalists are not willing to accept this. Louise Antony claims: “[...] there must be microphysical explanations of all non-basic properties and laws, including intentional ones, and... these properties and laws must be predictable from microstructural facts”. This must be so, if everything is physical. Antony belongs to that group of naturalistic philosophers who want to secure the scientific status of psychology. But is this an idle dream? Already Wittgenstein noted, when discussing the nature of psychology: “psychologists want to say: ‘There must be some law’ – although no law has been found.” Stoutland, repeating the Wittgensteinian point, has claimed that a modern metaphysician is considering how concepts must work in the metaphysical construction which he has built by and for himself. Hacker is more straightforward in his criticism of the trend of modern philosophy. He claims that the scientism which bedevils a notable part of contemporary philosophy has licensed also a scientific metaphysics. From Hacker’s perspective, Kim, Antony and others who insist that there must be such and such explanations are in the grip of scientism.

486 From Kim’s perspective, the positions of von Wright and Davidson are therefore hopelessly incoherent.
488 This could be likened to Searle’s “features”.
489 Antony, 1999, 38, emphasis mine.
490 Wittgenstein, 1966, 42.
491 Stoutland, 2006a.
492 Hacker, 2001. Also Davidson (1997c) has noted that a form of scientism is one of the obstacles standing in the way of mental realism.
The charge of incoherence against non-reductive physicalism is very common in the modern philosophy of mind. But, as we have seen, Davidson – a non-reductive physicalist – tries to vigorously defend physicalism. What is the place of mental properties in this picture? The position of non-reductive physicalism is usually understood in terms of irreducible properties, but the nature of these kinds of properties is far from clear. Searle’s definition does not clarify what a mental property is, and Kim’s definition merely states what mental properties are not. I think that the question about the nature of mental properties is important in itself, because in order to be able to discuss the problem of mental causation or the problem of emergent properties, a description of the nature of mental properties is required. Unfortunately, the contemporary discussion about mental properties is very obscure. Many reductive physicalists and non-reductive physicalists understand the nature of properties differently. It is easy and tempting to go along with the reductive physicalists who identify mental properties with physical properties. This is easy and tempting for the same reason: such a view fits best with the scientific understanding of the mind–body relationship. Moreover, in the discussions about mental properties, a paradigm example of a property is usually a property which in some clear sense is undeniably physical. It is therefore easy to use physical properties as examples of properties in terms of which mental properties should also be understood. This is a suggestive picture, which it is convenient to accept. But before this “easy” road is taken, we should perhaps ask whether the very question itself (i.e., are mental properties physical properties?) is distorted and a result of a linguistic confusion, or the result of accepting a false picture as a guideline for thinking.

2.3.1 The nature of mental properties

In this work I am not going to go deeply into a philosophical discussion about the nature of properties as such. I justify my decision not to discuss the general problem of properties or universals with the fact that Davidson and von Wright do not discuss the nature of this problem in any detail.\(^{493}\) The general discussion about the nature of properties is, or at least can be made to be, an extremely complex philosophical problem. I am not going to participate in this discussion. I will rest with von Wright’s and Davidson’s approach, as both in a sense stress the importance of common sense when approaching philosophical issues. Von Wright

\(^{493}\) In fact, in the posthumously released Truth and Predication (Davidson, 2005a), Davidson discusses in detail issues which relate to problems of predicates. However, these views do not really illuminate the question about the nature of properties.
states: “[…] the picture we create for ourselves of the object-quality relation… is founded on everyday experiences with macroscopic solid bodies and the changes which they may undergo”.

Without suggesting that this is Davidson’s view, I believe that what Davidson said about events can be said also about properties: “[…] the assumption… ontological and metaphysical… is one without which we cannot make sense of much of our most common talk.”

Since my purpose is to describe the positions of Davidson and von Wright, I do not want to become speculative by going beyond their views on this matter. I nevertheless find it quite absurd that the problem of mental properties is a central problem in the contemporary philosophy of mind, and yet only few philosophers bother to clarify or consider the question of what mental properties are. Those who bother to clarify the nature of mental properties are usually seen as being “overtly metaphysical” and are thereby excluded from the mainstream discussion in the philosophy of mind.

In the mainstream discussion it is assumed that one already knows what mental properties are. In this work I am willing to make, without argument, the assumption that physical properties are real. But how should mental properties be understood?

The question of whether mental properties are real may sound absurd. Of course they are real; they are at least as real as physical properties. I believe it is nevertheless easy to become confused with this question, because it is not clear what mental properties are. In the absence of this knowledge, philosophers use the term “mental property” without hesitation and formulate problems in terms of it. What is the common-sense understanding of mental properties? The reason one should consider this is that it is questionable whether the notions that refer to mental phenomena have, or could have, a “life of their own” completely detached from everyday usage. We recall that according to Wittgenstein’s conception of philosophy it was noted that “Psychological concepts are just everyday concepts.”

To many this sounds controversial. But where do we get the concept of “thinking”, in terms of which philosophical problems about the mind are formulated? This concept is, first and foremost, from our everyday language. Can this claim be “proven”? It is not easy to imagine what it would mean to prove it. Most humans use psychological concepts and have learned to use them in certain circumstances – this is uncontroversial. A child learns the use of psychological concepts as

---

494 Von Wright, 1998, 74. This resembles Davidson’s view that “physical concepts…are tied to the common sense notion of a physical object which has a location in space and time, which undergoes changes which are physical events…” (Davidson, 1964, 45).

495 Davidson, 1967a, 162.

496 For this kind of “overtly metaphysical” perspective from which also the problem of mental causation could be considered, see for example Lowe, 1989 and 2009.

(s)he becomes part of the human community. These are remarks about the natural history of humans. In their own ways Wittgenstein, von Wright and Davidson try to (re)turn our focus to what is already known. They are trying to highlight the mental or psychological facts obvious to all. But from the fact that psychological concepts are everyday concepts, it does not follow that the problems relating to them would not be immensely complex; perhaps it could be argued that the complexity is related to the fact that psychological concepts are so interwoven with the activities of human life. Wittgenstein tries to remind us of these complexities by discussing the manifold everyday circumstances in which psychological concepts are used.

Would scientific discoveries about the referents of mental concepts be discoveries about the referents of the mental concepts that are currently used? Mental concepts have conceptual connections to other common-sense notions and to the activities of everyday human life. These connections, as for example Bennett and Hacker have argued, are constitutive of the meaning of the current concepts. We should again remind ourselves of Wittgenstein’s lesson: “Pain has this position in our life; has these connexions; (That is to say: we only call ‘pain’ what has this position, these connexions)”.

Can a pain which does not have a certain familiar position in our conceptual scheme or certain familiar connexions be imagined? Perhaps. But what would be a reason to say that we are talking about pain as we know it? If the alleged pain did not have any of the conceptual connections that current pain has, how could it be identified as pain? Would it be rational to call such a phenomenon “pain”, i.e. would this expression make sense? Nothing would connect this new concept of pain to the concept of pain with which we are familiar. Could the new concept be part of our “form of life”? As Wittgenstein reminds, the concept of pain is characterized by its particular function in our lives.

The fact that pain has certain connections for other concepts and phenomena is empirical. It is a contingent fact that our concept of pain happens to have certain specific conceptual connections. It is also the case that people have different conceptions of pain. But all these ‘pains’ share features which are familiar from the language of everyday, and from everyday life. The joint conception, the features that are agreed upon, cannot be violated without changing the subject of discussion. It is not possible to force people to accept a new way of thinking about pain if it does not fit their way of life. It is difficult to see how

---

498 For an insightful discussion about why psychological concepts should not be seen as theoretical concepts, see Hacker, 2001a.
499 As Wittgenstein (1967, §64) reminds us: “[…] we need to remember that the process of thinking may be very various.” This theme is emphasized through Wittgenstein’s writings.
501 See, for example Wittgenstein, 1967, §101 - §103, §532.
scientific discoveries could be relevant for a deeper understanding of current mental concepts if these discoveries did not take as their starting point the concepts as they appear in everyday life. What would be the relevance of the discoveries for the current concepts? An approach which emphasizes the **primacy of the obvious** is common to those philosophers, Davidson and von Wright among them, who accept a mild form of behaviorism. \(^{502}\) The truth of this form of behaviorism seems to me something which cannot be denied. \(^{503}\) We *do* understand mental concepts in terms of behavior; the concepts have behavioral criteria. Von Wright’s last word on the subject was: “The criteria for the existence of individual mental phenomena are… behavioral and thus corporeal – and I fail to see how it *could* be otherwise.”\(^{504}\) For von Wright this is a conceptual truth; it is not an empirical question whether or not behavior is the criterion of the mental. This is emphasized by Wittgenstein with his famous claim that:”The inner is tied up with the outer logically, and not just empirically.”\(^{505}\) I find it interesting that this kind of Wittgensteinian–Davidsonian–von Wrightian behaviorism is being revived, of all philosophers, by Kim who writes: “[.]. .it seems to me that we cannot avoid thinking of intentional / cognitive states, like thought, belief and desire as supervenient on behavior and other observable physical facts. We must accept creatures that are behaviorally and functionally like us as creatures with mentality similar to ours.”\(^{506}\) According to Kim, the way we use language is among the reasons “for thinking that cognitive / intentional mental properties are closely tied, conceptually and semantically, to behavior.”\(^{507}\) Further mental states must ultimately be anchored, conceptually and epistemologically, in observable behavior. This kind of conclusion was one of the essential conclusions of von Wright’s last philosophical investigations.

My claim is that one crucial problem with the contemporary discussion about the nature of mental properties is that the primacy of the obvious has been ignored. What are the examples of mental properties of which we are directly aware? Let us try to consider this question without taking a stance on the philosophical question about the fundamental nature of properties. What do we mean by saying that there are such “things” as mental properties? The answer is not *obvious*; if this question is asked of a philosopher of mind, it is by no means certain that a clarifying answer is forthcoming. One “intuitively” plausible answer is that by claiming that there are mental properties, we are simply saying that mental states, or

---

\(^{502}\) This term is used by von Wright (2000) in characterizing the cornerstones of Wittgenstein’s later philosophy.

\(^{503}\) This claim will be defended in section 3.2.

\(^{504}\) Von Wright, UPe, 13.

\(^{505}\) Wittgenstein, 1992, 64e.

\(^{506}\) Kim, 2005, 166.

\(^{507}\) Kim, 2005, 167.
“phenomena”, are something which more than one individual can have. Mental states are thus something that can be exemplified by more than one individual. In the contemporary philosophical discussion, “being in pain” is often used as a paradigm example of a mental property. We can think of somebody’s being in a state of pain as exemplifying the property of “being in pain”, and similarly for any state that we take to be mental. Thus, being in a state S and exemplifying a property P are just two different ways of expressing the same thing. There is no essential difference between saying “I have pain” or “I am in pain” and “He [which refers to me] has [or ‘exemplifies’] the property of being in pain”. I see no reason why the terms mental property and mental state could not be used interchangeably.

It is commonplace to think that two persons can be in a state of pain. This being the case they can, and do, share a mental property, although it is unclear to me why this term should be brought into the discussion. Saying that two people are in pain is certainly part of the way we speak about mental phenomena, although the expression “mental property” is rarely used in colloquial language. In this sense, property talk is clearly the philosopher’s own creation. Mental properties are not needed in order to talk about mental phenomena in a meaningful and non-problematic manner. But if we assert that two people are in pain, are their pains the same? After all, your pain is yours, mine is mine. Since we do often say that two people are in pain, and we can fluently communicate with this kind of expression, it is best to consider what we mean by this statement. Here intuitions differ, and it is not clear how one could argue that a certain interpretation of the meaning of this statement is the correct one. If I ask myself, what do I mean when I say that Donald and Georg are in pain, where can I find the answer? I suggest that we again take a cue from Wittgenstein: “One ought to ask, not what images are or what happens when one imagines anything, but how the word ‘imagination’ is used.” One should not consider what mental properties are or what happens when someone is exemplifying a mental property, but how the expressions which refer to the relevant “mental properties” are used. Focusing on the first two questions is apt to lead to speculation, since we do not know what mental properties are or what happens in the brain when a person exemplifies this kind of property; there still remains an ontological puzzle about the mind. Therefore, one should focus on considering how psychological language is used. From this consideration, it can be seen what we mean when we use psychological terms, and it is doubtful whether any substantial puzzlement will remain when this use is clarified.

---

508 Wittgenstein, 1953, 370 §.
509 It could be argued that an ontological puzzlement remains. But if we think this, are we then merely misled by the thought that there must be a certain kind of answer?
If one considers the way mental language is used, it is possible to conclude, together with Davidson, that when we say that two people have the same thought, we mean that their states of mind are similar enough for each to be able to interpret the other and – when we take our own perspective in consideration – similar enough for us to interpret them.\textsuperscript{510} This interpretative approach to mental states is strongly based on the idea of the primacy of the obvious. Davidson claims: “We know what states of mind are like, and how they are correctly identified; they are just those states whose contents can be discovered in well-known ways.”\textsuperscript{511} It should be recognized that “well-known ways” should be taken literally. We all know how to attribute mental states to others; even very young children know how to do this. This again could be seen as note about the natural history of humans. We all know whether a person “exemplifies a mental property” at time t. The primary evidence for the attribution of “mental properties” is the behavior of others; children take this naturally into account without problematizing the connection between mental phenomena and behavior. What is essential in the views of von Wright and Davidson is the Wittgensteinian idea that the “nature” of mental concepts, thus of mental properties, is revealed through an analysis of how mental concepts are used. Wittgenstein did not use the term “mental property”, but he wrote about the problem in a way which is more illuminating than the discussion of contemporary philosophers of mind. The discussion about the status of non-reductive physicalism has largely ignored Wittgenstein’s observation that the use of psychological concepts is often confused. I believe it could be claimed, with some justice, that the most serious confusion of modern philosophy of mind in its critique of non-reductive physicalism is its attempt to understand and model the “nature of the mental” solely in terms of the physical. But if Wittgenstein is correct there is no such thing as the nature of the mental, and the mental cannot be understood as a metaphysical category. Incidentally, Wittgenstein’s claim that the mental is not a “metaphysical epithet” (logical category) bears a striking resemblance to Davidson’s latter claim that the mental is not an ontological category.\textsuperscript{512} Von Wright has the same view as Wittgenstein and Davidson: the mental is a conceptual category. Both Davidson and von Wright seem to follow Wittgenstein in his attempt to show the conceptual nature of the mental – especially the relation between mental phenomena and behavior.

When an interpretative approach towards mental states is taken, it is possible to see what is wrong with the very idea of a mental property as it is usually understood in the

\textsuperscript{510} Instead of “being in pain”, “having a thought t” is here used as an example of a mental property.

\textsuperscript{511} Davidson, 1988a, 40.

\textsuperscript{512} See Wittgenstein, 1992, 63e.
contemporary discussion; there simply are no such properties. Philosophers freely talk about mental properties as if it is perfectly clear what such properties are. Let us again use “being in pain” as an example. It is assumed that a specific property / concept, call it M, has enough stability so that the question of whether or not it is identical with a physical property – say, the firing of c-fibers – could be raised.\textsuperscript{513} Let us suppose that the human brain is observed from the outside and c-fiber activity is detected. Let us also suppose that the brains of person A and person B both have the property that c-fibers are firing, and that a more detailed, perhaps immensely complex, physical description of why the fibers are firing can be given. We can also assume that it is possible to describe a shared property of the brains at a micro-level “below” the firing of c-fibers.\textsuperscript{514} Firing of the c-fibers is a clear example of a physical property which both brains have. This can be said without problems. The brains of A and B have a property which can be observed objectively. It cannot be denied that there is a similarity between the two brains; perhaps their properties are even identical – in all relevant respects.

The problem is that the mental property of “being in pain” seems to be nothing like the physical property that can be observed intersubjectively. Whereas, based on intersubjective observations, it is possible to say with certainty that the c-fibers of two brains are firing, that in some sense of the word the same process is taking place in both brains, it is incorrect to say that two individuals who report that they suffer from pain have the same pain. We cannot conclude that the pains are even similar. The subjective element of “mental properties” is certainly something which distinguishes them from physical properties. Since this subjective aspect cannot be captured by anyone else other than the subject who is in pain it is, in my opinion, a grave mistake to think that pains could be grouped in such a way that the term mental property would be a clarifying description of “pain”. There are practically endless ways in which we can be in pain; that is, there are an enormous number of states in which we

\textsuperscript{513} Pain = c-fibers firing is commonly used by philosophers as an example of the identity between mental and physical properties. I don’t know whether this is or is even meant to be a neuroscientific fact, but it seems plausible that philosophers often use an expression like “the firing of c-fibers” without considering whether this expression captures any well-defined brain state at all. It is a mere placeholder in the discussion of philosophers. Kripke (1980) for example notes that this expression is used as shorthand for whatever neural mechanism science discovers about pain-processing in the brain. I claim that no matter what view we have about the vagueness of this expression, we will see that nothing could convince us that pain has to be identical with the firing of c-fibers. Of course, if we take the Putnamian–Malcolmian–von Wrightian line, we should conclude right from the start that the claim Pain = c-fibers firing is nonsense.\textsuperscript{514} I don’t know whether this is actually the case, but from the perspective of physicalism this, i.e. an explanation in terms of lower level entities, should be possible in principle. Certain forms of physicalism insist that such an explanation must be possible. Recall the view of Antony (1999, 38): “[…] there must be microphysical explanations of all non-basic properties and laws, including intentional ones, and… these properties and laws must be predictable from microstructural facts.”
take it as appropriate to use the term “pain” as a description of that state. To the extent that a headache, heartache – all of these are forms of pain – but do they have anything in common except our tendency to use the same term, ache, in each case?

One answer, which I think should be considered, is that what these pains have in common is the feeling of pain, and that is why each is called pain. The feeling, a sensation, is precisely what pain, as a mental phenomenon, is. Perhaps this feeling could be called an emotion. In the course of everyday life, pains certainly have more conceptual connections to emotions than to brain states. The centrality of the feeling is shown by the fact that, or so I assume, most of us would be very puzzled to hear that someone is in pain, but yet he feels nothing. Would we be equally puzzled to hear that someone is in pain but her c-fibers are not firing? The feeling – the “essence of pain” – is something that cannot be captured except by the subject whose feeling it is. How could you find out whether your pain feels the same as mine? If we cannot be sure that you feel the pain in the same way as I do, what would be our reason to think that it is the same pain that we both suffer?

Let us suppose that Donald has a headache and so does Georg. What is the reason to say that the headaches are similar mental phenomena? If they are not, it is misleading to refer to headache using the term mental property. Donald’s headache is a state of one individual; Georg’s headache is an ache of another individual. Donald’s headache is a property of him, but this says nothing about its relation to Georg’s headache, or to headaches in general. Can something general and illuminating be said about headaches, conceived as mental phenomena? I cannot see how, since they are always states of a certain individual. Even in the case of an individual like Donald, the question of whether his pain today is the same as tomorrow is difficult to answer. Maybe the pain of today feels almost similar to the pain of yesterday; yet there is an extremely slight, almost unnoticeable difference between them, which becomes “visible” for Donald only when he focuses on the pain through introspection. But can it be known that such self-observation does not change the feeling? Is Donald going through the same state twice, or are the pains two different states? If it is supposed that there is some difference between the states the latter seems to be more appropriate way to describe

---

515 The problematic nature of pain is shown also by the fact that psychologists, neuroscientists, doctors, cognitive scientists, philosophers and lay people all have very different views of what pain is. Also different sciences may have different definitions of what pain is. The International Association for the Study of Pain (IASP) defines pain as “[…] an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage” (See Merskey, 1986). Alternative definitions are certainly possible, and how could we decide which one is “right”? For example, is pleasant pain impossible? Why would IASAP’s definition have authority over the everyday concept?

516 For a discussion of pain without painfulness and painfulness without pain, see Grahek, 2007. For a recent general discussion about the philosophical problems surrounding pain, see Aydede, 2005.
what is happening; yet both states are called headaches, as if there were a “general headache” of which these two are instances and examples. It is obviously not false to say that both experiences are headaches; the confusion is to say that two headaches are examples of the same “mental property”.

In the philosophy of mind, during the last fifty years a lot of emphasis has been put on the claim that mental properties are multiply realizable and thus pain, for example, cannot be type-identical with a physical property. Whereas this may be true, the real problem, it seems to me, is that it is not possible to say whether pains, as mental phenomena, have any shared features with each other except the subjective feelings that connect them, which cannot be captured in any objective way. But if there are no general mental properties, one of the main problems posed for non-reductive physicalism vanishes. If mental properties are being modeled by using a physical property as an example, a dead-end is quickly reached because physical properties have no subjective aspect. This aspect is, however, the essence of a “mental property”: this aspect could group pains together, it would show their similarity to each other, if the aspect could be brought under objective analysis. But an objective analysis of a subjective aspect is a contradiction in terms. If sense could be made of the claim that Georg’s pain feels (and therefore is) the same as Donald’s, it would be possible to say that they shared a mental property, since there would be something which both of them would have; the same feeling. Lacking this alternative, it makes no sense to say that they share a mental property. To insist that they must feel the same pain because they are in the same physical state is obviously wildly question-begging. To suggest that there is no such thing as the subjective feeling of pain is absurd. This view generalizes also to other mental phenomena than sensations. The identity conditions for “having the same belief” are even more obscure than the identity conditions for sensations, because whereas, in the case of sensations, there is something which could be similar – the feeling – we do not know what to even look for in the case of beliefs.

In its attempt to bridge the gap between the mental and the physical, contemporary philosophy of mind has largely ignored the discussion about the nature of mental concepts. A lesson from Wittgenstein is again appropriate in order to see what has been ignored. He writes: “We are not at all prepared for the task of describing the use of e.g. the word ‘to think’… the naïve idea that one forms of it does not correspond to reality at all.” The concepts have been removed from everyday use and as a result a distorting view about the

---

517 Wittgenstein, 1953, 111, my emphasis.
nature of mental “properties” has occurred. Mental concepts are from everyday language. How could they remain the same concepts once removed from this context? If removed from such a context, the result is a naïve view such as the contemporary suggestion that there are mental “properties”. Wittgenstein’s view contains also an insight about the source of the contemporary problem. He says:

How does the philosophical problem about mental processes and states… arise? The first step is the one that altogether escapes notice. We talk of states and processes and leave their nature undecided. Sometime perhaps we shall know more about them – we think. But that is just what commits us to a particular way of looking at the matter. For we have a definite concept of what it means to learn to know a process better.

That naturalistic philosophy of mind is committed to a particular view of mental properties is obvious. Naturalism has a definite concept of what it means to learn to know a mental process better. Naturalists claim that if mentality is real, it must be something other than what it appears to be. As Wittgenstein thought, this assumption is one of the biggest mistakes that can be made in philosophy. Those who are sympathetic to “Wittgensteinian approach” in the philosophy of mind have emphasized that Wittgenstein’s lesson should not be forgotten. Anthony Kenny suggests:

[…] some of the philosophical gains we owe to Wittgenstein seem in danger of being lost… his contribution has been neglected because more and more philosophers, especially in the United States, have attempted to model their studies on the pattern of a rigorously scientific discipline, mimicking the type of precision characteristic of mathematics, and holding up… an abstract system of artificial intelligence as the goal of philosophy of mind.

This is a very perspicuous comment. It is unfortunate that the naturalistic trend silences Wittgensteinian insights without arguments, but it is even more unfortunate that this trend itself creates philosophical pseudo-problems, like the ‘problem’ about the nature of ‘mental properties’. It is the first step, as suggested by a physicalistic metaphysics, which leads one astray. Malcolm claims, in a Wittgensteinian spirit, that:

Our assumption that there is a nature of… thinking… to be found out, to be identified or explained, either by philosophy or by science, is the worst mistake we make in the philosophy of mind…. We start

---

518 Strictly speaking there cannot be any “false” or “right” view about the nature of mental properties because the only measuring stick is the conformity or non-conformity of this expression with the way that language is used. What can be said is that the more the use of this expression distorts the way it is actually used in the everyday language, the more “false” it is.
519 Wittgenstein, 1957, § 308.
520 Kenny, 1984, vii-viii.
off by saying to ourselves: ‘What is remembering? What is the process of remembering? And already we have gone wrong!’\textsuperscript{521}

More recently a similar observation has been made by Stoutland:

\begin{quote}
It is taken for granted that various events, processes, and states are there to be investigated and that our primary philosophical task is to construct a theory about their nature and relationships. But this neglects the vital task of considering how we understand them as we do…\textsuperscript{522}
\end{quote}

It is a telling fact about contemporary philosophy that Stoutland must emphasize the same point that Malcolm did twenty-five years earlier, as did Wittgenstein twenty-five years before him. There is truth in Hacker’s observation that contemporary philosophy of mind, far from showing Wittgenstein’s arguments to be invalid, has simply chosen to ignore these arguments.\textsuperscript{523} I believe that the reason for this ignorance has one clear source. I suggested in chapter one that the nature of philosophy is changing. Many modern philosophers think that philosophy should construct explanatory theories. When this task is undertaken, underlying conceptual problems remain unsolved and keep on surfacing. Language, the source of such problems, has stayed the same, thereby giving rise to the same problems again and again. A concrete example of this is the debate between Hacker / Bennett and Searle / Dennett. The debate shows that no real progress has been made: the conceptual problems are basically the same as they were in the times of Wittgenstein, whose prediction the debate seems to confirm.

I think we should conclude that the expression \textit{property} should be understood differently in the mental and physical cases. Because of the subjective character of the concept involved, it is \textit{never} possible to say whether an instance of a mental “property” is identical to another instance – even though we refer to both instances, with the same name and know that these two instances share some features. If an uncontroversial physical property is the paradigm example of a property, then there are no such things as mental properties. The fact that the term “property” is used in both cases creates a distorting misunderstanding; that the contemporary discussion about the mind–body relationship has been carried on in terms of mental properties is therefore unfortunate.

\textsuperscript{521} Malcolm, 1970, 29.
\textsuperscript{522} Stoutland, 2005, 134.
\textsuperscript{523} Hacker, 2001. Von Wright has stressed the same point.
2.3.2 Why mental properties create a problem for non-reductive physicalism

The conclusion of the previous section was that there are no mental properties in the sense in which most (physicalistic) philosophers talk about them. Their talk of mental phenomena resembles too much their talk of physical phenomena. In his time, Wittgenstein tried to warn that mind should not be seen as standing for a substantive ‘thing’. Modeling mental phenomena by using physical phenomena examples creates the misleading picture that there are such things as mental properties. Pain is claimed to be a mental property. Yet, we do not know how to fill the blank in the statement: “To be in a state of pain is…” Kim, for example, claims that: “we know… that pain occurs only because a certain neural state, call it $\Psi$, occurs.” Kim does not refer to neurophysiological evidence, but I assume that he has in mind something like “Pain = c-fiber firing” when he says that we know that pain occurs because $\Psi$ occurs. This view, that pain occurs because a “certain neural state” occurs, is a naïve one even in terms of the available physicalistic evidence, because during a painful stimulation a PET-scan allegedly shows brain activity in the sensory and motor cortices, premotor cortices, parts of the parietal and frontal cortices, cingulated cortex, insula, and occipital cortex. Pain also projects to the following subcortical structures: thalamus, putamen, caudate nucleus, hypothalamus, amygdala, periaqueductal grey matter, hippocampus, red nucleus, pulvinar, and vermis of the cerebellum. Pain, as a physical phenomenon, seems to be a more complex issue than what physicalists in their simple picture suggest. Such a simplification may lead to many philosophical mistakes about what pains, or more generally mental phenomena, are. Perhaps it could be argued that, also in terms of neuroscience and not just mentally speaking, that pain is a global phenomenon.

In order to think of pain as a property, we should be able to say what pain is and what pains have in common. I may describe on each occasion what my being in pain means, how I see the situation on certain occasion – but this cannot be generalized so that it would be a clarifying description of the property that “being in pain” is supposed to be. Perhaps pain simply is a phenomenon the nature of which cannot be fixed, and perhaps what von Wright wrote about values could be said about pains as well: pain is something which we can understand but which cannot be codified into truths. Can it be proven that the nature of mental phenomena or concepts cannot be clearly defined? This seems unlikely. But what would be the reason to think that pain does have a fixed nature? There are various definitions of pain –

---

524 Kim, 2004, 135, emphasis mine.
525 See Wall, 1996.
for example the one given by IASP – and these definitions are not immune to counterexamples. Even if it is accepted that pain is, by definition, an “unpleasant sensory and emotional experience associated with actual or potential tissue damage”, or is described in terms of such damage, this does not help at all in the consideration whether the headache of Donald feels the same as the headache of Georg. The definition cannot help to solve the question whether Donald and Georg have the same pain. In one clear sense they certainly do not have the same pain: Donald’s pain “belongs” to Donald, Georg’s pain “belongs” to Georg. But ignoring this aspect, is it the same pain in both cases? Is this a meaningful question? Is there a way to tell whether the pains are the same? The definition of IASP refers to “unpleasant sensory and emotional experiences” and thus, or so it seems to me, the crucial question is precisely whether these experiences are the same in the case of the two pains. Pain has been used here as an example, but the problems are even more obvious with more complex mental phenomena like jealousy, desire, hate, or love. There is absolutely no reason to think that a satisfying definition of jealousy, a definition which would apply to every person who is jealous and to none who is not, could be found. Donald shows certain behavioral signs which suggests for us that he is jealous, whereas Georg admits that he is feeling extremely jealous. The absurdity of the question “Who is really (or more) jealous, Donald or Georg?” shows how desperate task it would be to try to form a satisfactory definition of jealousy.

Despite the obscurities involved with the expression “mental property”, it is used non-problematically in the current philosophical discussion. This means, I believe, that those involved in the discussion think that they understand what this expression means and how it should be used. It is claimed that mental properties create a serious problem for a non-reductive physicalist by forcing him to accept “property dualism”. Kim writes: “[a] nonreductive physicalist believes that there are events in her ontology that have mental properties (e.g. being a pain, being a belief that snow is cold, etc.).” This seems correct, although it is unclear whether many non-reductivists would accept the claim that events have mental properties. It is also assumed by Kim that a non-reductive physicalist is a realist about (mental) properties. Let us accept this for the sake of argument. What is problematic is Kim’s view of the nature of mental properties. He says that in formulating the problem of mental properties: “nothing will depend on precise general definitions of ‘physical’ and

---

526 Sometimes definitions help. Water = H₂O does help to decide whether a glass contains water.
527 Kim, 1989b, 279.
528 See, for example, Kim, 1993b.
‘mental.’ But how could this be the case? If there are no mental properties of the sort Kim thinks exist, then surely the whole problem of their relation to physical properties, as formulated by Kim, disappears. In order to discuss the problem, we must know what a mental property is. An assertion that “being in pain” is a mental property is not enough.

The misunderstanding of the nature of mental properties is the main reason for the confusing way in which the accusation of property dualism is put forward. Paul Churchland describes the idea of property dualism as follows:

[...] the brain has a special set of properties possessed by no other kind of physical object. It is these special properties that are nonphysical: hence the term property dualism. The properties in question are the ones you would expect; the property of having a pain, of having a sensation of red, of thinking that...

It should be noted that whereas Kim talks about events that have a property of being a pain, Churchland talks about brains having the property of having a pain. It is not obvious that a non-reductive physicalist would be committed to either one of these theses. He could argue, for example, that events or brains do not have pains, only persons do.

The “dualistic” claim that the brain has properties that no other physical object has sounds right. As far as we know, the capability, for example, of being self-conscious is a property that only the human brain possesses. This cannot be the dramatic claim of property dualism. It is claimed that according to property dualism “some material entities can have properties that are essentially different from those we normally regard as physical properties.” Given the variety of physical properties, it is difficult to understand what is the key characteristic of these properties is that marks them as “essentially different” from the others. It should also be asked what the nature of the properties which we “normally” regard as physical is: are “normal” physical properties those which are evident in everyday life or those studied by physics? The subjective aspect of mental properties could perhaps be a feature which marks an essential distinction between them and physical properties. But is there anything “dualistic” in this claim? No; to say that the brain has properties which no other physical object has is not a dramatic claim. This is especially so if one thinks that mental

---

529 Kim, 1993b, 340.
530 Instead of calling this approach a misunderstanding, it is perhaps better to say that only a specific understanding about the nature of mental properties is what opens up the possibility for the problem of property dualism.
531 Churchland, 1988, 10.
532 This is a controversial claim since many think that also some other animals are capable of self-consciousness. Let us ignore this possibility for the sake of argument.
533 Svensson, 1994, 97, my emphasis.
properties are a result of lower level physical properties, i.e. if it is thought that the former supervene on the latter.

The claim that sounds implausible from the physicalistic viewpoint must be the claim that the “property of having a pain” is a non-physical property. But what does non-physical mean here? It means that these properties do not appear in basic physics or cannot be reduced to the properties of basic physics – or, more broadly, that these properties cannot be explained solely in terms of the concepts of the physical sciences.\(^{534}\) This, most likely, is the essential difference between mental and physical properties. Everything about the latter can, in principle, be explained in scientific terms, whereas this is not the case with respect to the former. For a property to count as physical, it should be such that it can be reduced, “in a broad sense”, to fundamental physical properties. But whether or not a non-reducible property should count as non-physical seems to be, above all else, a purely definitional or terminological matter. It can be argued that irreducible properties are non-physical in the sense that they are not reducible, but nothing substantial or interesting, as far as ontology is concerned, follows from this. In section 2.3 it was noted that Searle’s claim that “the irreducibility of the mental to the physical is a sufficient proof of the distinctness of the mental” is obscure and suspect if the “distinctness” in question refers to ontological categories. What is the rationale for believing that everything physical, everything that exists, could be explained in physical terms?\(^{535}\) Faith that science will advance to a stage where this may happen is the rationale. But this is mere speculation, and as the results of quantum physics show, it may be that the nature of reality is such that “absolute” physical explanations are simply not possible. This would not be a result of the inadequacies of our epistemological capacities, but rather a feature of reality. Whatever the situation turns out to be, it should be acknowledged that physicalism is not really a scientific view but a metaphysical doctrine.

The new-wave reductionist John Bickle claims that property dualism remains a form of dualism since “[…] it denies that even a matured physical science could exhaustively explain the essence of the mental.”\(^{536}\) Some dualists certainly make this kind of claim, but the same is true of physicalists as diverse as Kim, Mcegin, Chomsky and Nagel. Given the historical load of the term “dualism” and its close connection to Descartes’ view about essentially different substances, the term is quite inappropriately associated with non-reductive physicalism. To suppose that a matured physical science could explain the “essence

\(^{534}\) For these formulations see, Kim, 2003 and Churchland, 1988.
\(^{535}\) For a view questioning this rationale see Meixner, 2004, 2005, 2006.
\(^{536}\) Bickle, 1998, 7.
of the mental” is only to show allegiance for the view that reductive physicalism is true. To insist that physical science must be able to explain the essence of the mental is merely to state a dogma of physicalism. It is almost like some physicalists do not want reality to be such that a part of it could remain beyond explanation in physical terms. Given that the “physical essence of the mental” has surely not been explained by a physical science, opposed opinions about the irreducibility or reducibility of the mental stand on an equal footing. But, terminology aside, does a non-reductive physicalist really accept the view that “mental phenomena are distinct from neural phenomena”? In my opinion this is more like an interpretation of a non-reductivist view than a position which, for example, Davidson and von Wright would accept. We should ask what the claim that mental properties are distinct from neural phenomena means, according to those who make this interpretation. If it means that mental properties will not reduce to fundamental physical properties although everything is composed of physical particles, then the dualism of which non-reductive physicalists are accused of is at best an attenuated form of dualism, as for example Kim admits. So, instead of making the accusation, as Bickle does, that non-reductive physicalism is essentially a form of dualism, it could be noted, as Kim does, that this position seems to accept a form of very weak dualism in a way similar to biological naturalism.

Some physicalists are eager to claim that higher level physical properties will reduce to basic physics, but the question of whether this will actually happen, or whether such a reduction is possible even in principle, is pressing. It is unclear whether a matured physical science is capable of explaining even the “essence” of matter. If higher level properties in general may turn out to be irreducible to basic physics, there is no rationale for claiming that the irreducibility of mental makes a non-reductive position any more “dualistic” than the position of physicalism itself. Kim, among others, often runs together the terms “non-reductive physicalism” and “property dualism”. This is misleading, because there are also forms of property dualism which are genuinely dualistic. I would thus agree with Pereboom and Kornblith when they suggest: “Under no classification is the anti-reductionist’s position a kind of dualism... it is pluralism at one ontological level and a monism at the most

---

537 Bickle’s formulation raises further questions: What is an “exhaustive explanation”? What about the “essence of the mental” is still unexplained?
538 Recall Smart’s (1959, 142) “confession of faith”: “That everything should be explicable in terms of physics... except the occurrence of sensations seems to me to be frankly unbelievable.”
539 Kim, 2005, 153. See also Antony, 1999.
540 Kim, 1993b.
541 This problem is raised for example by Chomsky, 2000.
The real issue is not whether a certain position should be labeled as dualistic or not. The important question is whether non-reductive physicalists accept the kind of distinction between mental and physical properties that is suggested by the critics. Even this question would be uninteresting if it were not repeatedly claimed that the relation between mental and physical properties is an important problem in the contemporary philosophy of mind requiring a solution, and that a non-reductive physicalist cannot provide it. “Property dualism” often refers to a position that a non-reductive physicalist, like Davidson or von Wright, would not accept. In this merely exegetical sense, the correctness of certain labels of certain philosophical positions is of some interest.

In the following I shall tentatively (and for the sake of argument) assume that the notion of “mental property” makes sense. Let us assume a certain kind of understanding of mental properties in order to better evaluate the problem that is posed for non-reductive physicalism in general and for Davidson’s position in particular. The problem could also be formulated without the confusing term “mental property”, and this would dispel the question about the coherence or incoherence of property dualism. As Kim notes: “the substantive question that we are asking, or should be asking, is whether or not things like belief, desire, emotion and consciousness are reducible to neural, biological, and physicochemical properties and processes.” Indeed, this question, i.e. whether “mental things” can be reduced to physical phenomena, is the interesting one. But this question of whether a “mental thing” is reducible to a physical one sounds like the question of whether token-physicalism is true. If this is the substantial question, the answer of non-reductive physicalism is clear. Davidson argues that belief and desire are token-reducible and von Wright thinks that token-identity is trivially true and local reduction is possible. Davidson was in fact one of the first philosophers to argue that the answer to Kim’s “substantive question” has to be affirmative. Given that Davidson and von Wright answer Kim’s substantive question affirmatively, how can they be described as dualists or non-reductive physicalists? If the question is answered affirmatively, perhaps the status of non-reductive physicalism depends on the question of whether “things” like belief and emotion are de facto reducible. I will consider this question in section 2.4.

We should note that if, as Kim claims, the substantive question is whether a thing like consciousness is reducible to neural processes, then the answer also among reductive physicalists may be a reserved no. There are hardcore reductive physicalists who would not shy away from the claim that consciousness is going to reduce to neural processes. This claim

---

543 Kim, 2005, 160.
is intertwined with the problems relating to the questions of what consciousness is and what its reducibility to something else would mean. These are not just empirical questions but conceptual ones. As von Wright has claimed: “no future ‘brain research’ will – contrary to what many enthusiasts now seem to expect – eventually solve ‘the riddle of consciousness’. This… is a philosophic muddle which no findings of a scientific nature will ever clarify.”

If this view about the problem of consciousness is accepted, an empirical solution just isn’t the right kind of answer. This view cannot be ignored by insisting that the problem of consciousness is an empirical problem; there has to be room to think about the problem in characteristically different ways. Excluding the most enthusiastic reductionists, the consensus view among naturalists about the actual reducibility of consciousness may in fact resemble the view of non-reductive physicalists: among others Fodor, Chomsky and McGinn have expressed their different doubts about the reducibility of consciousness. Kim, who is certainly a reductionist, thinks now that the mind will not completely reduce to the physical. The current view seems to be that mind will not smoothly reduce to matter. This being the case, the old mind–body problem is still a problem.

If the substantial question is, or should be, formulated in the way that Kim does, then it is unclear whether there is any problem of mental properties. If token-identity is accepted, it seems that the question about type-identity is, ontologically speaking, a side issue. Smart, a recognized type–type identity theorist, perspicuously notes:

How far token-token identity theories can be extended in the type–type direction is a matter for further empirical conjecture and investigation. On the main ontological issues the differences between the two types of mind-brain identity theory do not seem to be crucial.

It is easy to agree with the view that the strength of type–type relations is a matter of empirical investigation and, as far as ontological issues go, token-identity theories and type–type identity theories are not crucially different. But it should nevertheless be noticed that the important question is not how well mental types correspond to physical types. The essential question is not whether type-identity or token-identity best captures the relation between the mental and the physical. The prior problem is whether sense can be made of the idea of a mental type.

If it is agreed that there are mental properties in the sense that there are mental states in which two people can be, the problem becomes: does a similarity of mental states imply a

544 Von Wright, 1989b, 30.
545 Kim, 2005.
546 Smart, 1989, 134, my emphasis.
similarity of physical states? This, in my view, is the best way to formulate intelligibly the question of whether the properties in question might be identical. Let us assume that pain is a well-defined mental property. Is it the case that when people are in pain they are always in a physical state P? To this question type–type identity theorists would answer affirmatively, but this theory is not widely supported anymore. Why not? As I briefly noted in section 2.1.1 there were two major reasons why type–type theory started to seem implausible. First was the advent of a functionalist view about the mind and the argument that mental properties are multiply realizable. According to this view, mental properties are defined by the functional role that they have in the behavior of their bearer(s). A second argument against type–type theories was Davidson’s view that there cannot be strict psycho-physical laws. But do these arguments really show that it could not be, or cannot be, the case that whenever Donald is in pain he is in a physical state P, or that whenever Georg is in pain he is in the same state P as Donald? They obviously do not show that it could not be the case. What could show that? It is perhaps possible to show empirically that Donald is not in state P, although he is in pain. But after showing this we are faced with the question of whether the pain, as a mental phenomenon, at t₁, is type-identical to the pain, as a mental phenomenon, at time t. If it is not, then it is not a mystery why the second of these pains fails to be identical with P although the first is identical with P. Suppose that Donald were to insist that his pain at time t₁ felt exactly the same as the pain he had felt at time t. How could we be sure that he is not mistaken? Perhaps there might be such a slight difference in the intensity of two pains that it would be almost impossible to detect it, and Donald would thus ignore the difference. Yet, this possibility would mean that the pain, as a mental phenomenon, at time t₁, was not exactly the same as the pain at time t. So the fact that a pain at time t₁ fails to be identical with a physical state, or let us say with property P, does not mean that pain could not be identical with P. which was identical with the pain at time t.

Do functionalism and AM show that it is not de facto the case that when Georg and Donald share a mental characteristic they also share a physical characteristic? Of course not, how could they show this? Suppose that we observe, through familiar means, that Donald is jealous and so is Georg. Whether or not “being jealous” is a property which could be physically multiply realized, the mere possibility of it being so realizable does not settle the question of whether Donald and Georg in fact share a physical property on which their being jealous supervenes, or which is identical with their being jealous. Whether or not they share

---

547 An interesting exception is Polger, 2004.
such a *physical* property is an empirical question. Whether or not mental properties are in fact multiply realizable is an open question. Currently the exact answer is unknown. The alleged multiple realizability of mental properties is sometimes defended by imagining non-human creatures which have such a strange physical constitution that their being jealous cannot be identical to a physical property P because they do not have properties like P at all. But it is utterly implausible to think that this argument would show that Donald and Georg do not, or could not, or *must* not, have an identical physical property P when they are being jealous. It is interesting that the multiple realizability argument has gained so much support, given the fact that it was originally formulated in terms of intuitions about the mental lives of non-human animals or even extraterrestrial creatures. Perhaps it could be claimed that the point of the multiple realizability argument is philosophical. It is not the case that a mental property M *cannot* be realized by various physical properties. M would not thus be identical with P. But, the argument does not show, and cannot show, that it could not be the case that a specific M *must* be realized by P.  

What about Davidson’s argument, according to which there are no psycho-physical laws? If Davidson is correct it can never be shown or proven that a certain mental property is identical with a physical property P. Tight connections between the mental and physical *descriptions* in question are impossible. But, or so the naturalists would claim, the question about the identity of properties is an ontological one. Does the absence of psycho-physical laws show that mental properties are not physical properties? No, what the argument shows is that it is not possible to *determine* the physical properties that are identical with the mental properties, assuming that they indeed *are* identical with physical properties. Does Davidson’s argument show that it is not possible that when Georg is in pain and Donald is in pain they both have a physical property P, and that they have pain because P exists? Does the argument show that this could not be, or is not, or must not be the case? No, it leaves these questions open.  

As we have seen, the contemporary discussion about mental properties in the context of non-reductive physicalism includes the claim that non-reductive physicalists are property dualists. But how could a physicalist be a dualist of any kind? The accusation, of course, is that she can’t, and that this is precisely what is wrong with non-reductive physicalism and its non-physical properties. But in what sense are mental properties non-physical? According to a

---

548 This discussion is immediately confronted with confusions about modalities. These cannot be completely ignored when discussing the status of multiple realizability argument, but they can be ignored here when discussing the question of whether mental properties are *de facto* identical with physical properties.
non-reductive physicalist, the non-physicality of mental properties means that these properties cannot be found among the properties of basic physics, and that they cannot be reduced to such properties. But this does not mean that mental properties are non-physical in any substantial sense, it does not mean that mental properties inhere in a non-physical substance or that they do not function according to the laws of nature. Donald’s becoming jealous, his acquiring this mental property could, in principle, be explained in terms of physical properties. A non-reductive physicalist does not need to reject the possibility that such an explanation could be given. Given a commitment to O-physicalism, why think otherwise? Property dualism is a badly chosen term to describe the non-reductive physicalist’s view of the relationship between mental and physical properties. Those who accuse non-reductive physicalists of property dualism have chosen a confusing and misleading expression to describe the position – and most interpretations of, for example, Davidson’s position therefore simply do not accurately capture the spirit of his work.

I claim that we should conclude (in the context of this discussion) the following: we should accept that there are mental properties and by this I mean mental phenomena that people can and do “share”. The contemporary discussion in the philosophy of mind is largely based around the assumption that mental properties are something real. Let us accept this for the sake of argument. This being the case, the central question has been what their relation to physical properties is. A philosopher committed to O-physicalism has only one answer: in some sense they are physical properties. Davidson, although admitting that everything can be described in the language of physics, however denies that mental properties are physical. Is it because: (a) he rejects properties altogether; (b) he rejects mental properties; or (c) he is a substance dualist? Since at this point we have no reason to believe in (c), let us consider the alternatives (a) and (b). I now turn to a discussion of this question in order to tease out one way to understand the nature of mental properties in the context of physicalism.

2.3.3 Davidson on properties

The purpose of this section is to come to an understanding of what Davidson’s view of properties is and what the reasons for it are. I defend Davidson’s position against those critics who describe it as a form of property dualism and thus defend my interpretation, given in section 2.1.3, that Davidson argues for robust physicalism.

According to AM, every particular can be described in the language of physics but mental properties are not physical properties. How should the latter claim be understood?
Doesn’t a dualism of properties imply substantial dualism rather than conceptual dualism, which Davidson certainly accepts? For example, in his analysis of dualism, Howard Robinson makes a distinction between predicate and property dualism, arguing that the latter is, ontologically speaking, a stronger position.\textsuperscript{549} He refers to Davidson as an example of a predicate dualist but as we have seen there is a strong tendency to describe Davidson as a property dualist as well.\textsuperscript{550} Davidson is often described as supporting a view which is opposed to robust physicalism. Ilkka Niiniluoto, for example, uses the term emergent materialism to describe Davidson’s position.\textsuperscript{551} This is interesting given the fact that, unlike von Wright, Davidson did not want to describe his view of the mental in terms of emergence. Niiniluoto’s use of the term is thus at least misleading if not incorrect with respect to Davidson’s own understanding of the nature of his position. As an exegetical remark it can be noted that Von Wright’s position, on the other hand, could with some justice earn the label “emergentism”. As I have suggested, this could be done as a consequence of his strict physicalism. Davidson’s position, however, is certainly closer to a view that: “we can have ontological mind–brain identity without mind–brain reduction... the mystery occurs... because of epistemological constraints on our abilities to analyze complex systems”\textsuperscript{552} than to a view according to which “there exists two fundamentally different kinds of properties....”\textsuperscript{553} It is clear, however, that Davidson is a property dualist in the sense described by Kim, namely: “[mental properties] are nonphysical in the sense that they do not appear in our basic physics and... they are not reducible to the properties dealt with in physics.”\textsuperscript{554} We have seen that this is one standard way that contemporary property dualism is understood, but in my view it is doubtful whether this kind of position should earn the label of dualism.

Ernest Sosa has noted: “Either AM [anomalous monism] accepts properties or it does not. If it does, then is not AM committed essentially to property dualism?”\textsuperscript{555} I think this is precisely the case given Davidson’s view about the status of mental properties. If Davidson is a realist about properties and if mental properties are not reducible to physical properties, then this implies a more substantial dualism than the conceptual dualism to which Davidson is committed on his own admission. On the other hand if Davidson is an antirealist about properties, this could open the way to an austere form of physicalism. Sosa makes an

\textsuperscript{549} Robinson, 2003.
\textsuperscript{550} Antony, 1999, for example describes Davidson as a property dualist.
\textsuperscript{551} Niiniluoto, 1990.
\textsuperscript{552} Uttal, 2004, 216.
\textsuperscript{553} Uttal, 2004, 215.
\textsuperscript{554} Kim, 2003, 114.
\textsuperscript{555} Sosa, 1993, 48.
interesting remark when he notes that if Davidson rejects properties, then “reality may be in itself homogeneously physical, since all particulars, all particular token events, would be physical. This seems a more radical and interesting, a truly monistic physicalism.” On this reading Davidson would be close to an eliminativist. This interpretation would agree with my claim that Davidson argues for robust physicalism. But at the same time it is difficult to discuss these problems and Davidson’s position without the threat of conceptual confusion. As we have seen, it is questionable whether the irreducibility to physics is enough to render properties non-physical in more than a merely terminological or definitional sense. On the other hand, there are views, like genuine emergentism, panpsychism, idealism and substantial forms of dualism which are clearly something stronger than the property or predicate dualisms of non-reductive physicalism. Where on this continuum Davidson’s view falls is a difficult question, which depends on one’s view, or on one’s intuitions as one could say, of what the irreducibility to physics entails. If a principled irreducibility to physics is interpreted as implying a form of dualism, then the term “emergent materialism” is perhaps not completely out of place as a description of Davidson’s position. On the other hand, at least some of the classical emergentists thought that emergent properties are something novel and ontologically over and above physical properties. The spirit of Davidson’s philosophy of mind does not go well with this kind of view. This being said, it should also be considered that if the irreducibility implies that the essence of the mental will remain mysterious forever, then non-reductive physicalism is perhaps best described as a form of dualism, at least in spirit if not in practice. The essence of the mental would remain mysterious, but a non-reductive physicalist could still firmly argue against the immortality of the soul or against other views which seem to be in conflict with a scientific understanding of the world (but which are important or essential aspects of certain versions of dualism).

Davidson claims that mental properties are not physical properties. If he is a realist about mental properties, then the claim about the distinctiveness or non-identity of mental properties with physical properties would amount to a form of strong property dualism, perhaps resembling something like what is proposed by Chalmers. According to him:

[...] property dualism... involves properties of an individual that are not entailed by the physical properties of that individual. Consciousness is a feature of the world over and above the physical features of the world.... there are properties of individuals in this world... that are ontologically independent of physical properties... [the] property dualism that I advocate involves fundamentally new features of the world.  

556 Sosa, 1993, 49.  
557 Chalmers, 1996, 125.
If and when property dualism is attributed to Davidson, we should consider how strongly such a dualism should be understood. The question has wider importance; it is one of the central questions to be addressed in the attempt to formulate a non-reductive view of the mind. If Davidson did subscribe, for example, to Chalmers’ version of property dualism, then the nature of his physicalism would be questionable. I shall not take a stance on the question of how the position of Chalmers is best understood, but the expression that mental properties are “fundamentally new features of the world” which are “ontologically independent” can be thought of as a very strong form of property dualism. Here again the question about the “distinctiveness” of mental properties is essential; ontological independence could be taken to mean the true distinctiveness of mental properties. Chalmers wants to distinguish his position as being a stronger view than the “dualistic” view, according to which mental properties are not properties invoked by physics or reducible to such. This weaker sort of property dualism is the one which Kim and Churchland attribute to non-reductive physicalists. The position of Chalmers, on the other hand, seems to be the kind of position of which certain physicalists talk about when they hurl their accusations of genuine property dualism. This position would be *dualistic enough* for such accusations to stick.

As an answer to the question how of Davidson sees the nature of properties, it could be argued that given Davidson’s Tarskian and Quinean influences, a natural conclusion is that Davidson is a nominalist and denies the existence of properties altogether. But whereas Davidson is skeptical about properties, he claims that he is not a nominalist. Strangely enough, many critics do not notice or care about this and still continue to describe him as a nominalist. Stoljar, for example, claims that Davidson’s position “officially eschews properties outright in favor of a nominalist ontology of events”\(^\text{560}\). The reason why Davidson refuses to accept properties as *concrete* entities in his ontology is his conviction that everything involving properties can be explained by treating them as *abstract*. I think it is not clear how the terms “abstract” and “concrete” should be understood here. How does the claim about the abstract nature of properties fit with some other of Davidson’s claims, like the following: “The relevant laws [which explain why a piece of chalk broke] *have to do with the microscopic properties* of this particular piece of chalk…”, \(^\text{561}\) or that “laws deal with types of events, and hence with particular events only as they have the properties that earn them

\(^{558}\) For a brief but useful clarification of Chalmers’ position see Chrucky, 1998.

\(^{559}\) Davidson, 1997c.

\(^{560}\) Stoljar, 2008, 276. For a similar view see Campbell, 2006.

\(^{561}\) Davidson, 1964, 48.
membership in a type”? Isn’t there a distinction made between predicates and properties when Davidson claims: “[…] concepts classify in terms of properties that may or may not stir the senses at the moment”? How can predicates classify in terms of properties if there are no such things as concrete properties, or how can properties “stir the senses” if they are abstract? Let us suppose that a piece of chalk has such and such microscopic properties and we refer to them in order to explain why the chalk broke. How can we give this kind of explanation if the property in question is not concrete? David Armstrong has asked: “Must there not be something quite specific about the things which allows, indeed ensures that the[se] predicates apply? The predicates require ontological correlates.” Von Wright makes the same point: “[a] predicate names something which has ‘ontological status’, viz. a quality”. Armstrong argues that even simple explanations become obscure if predicates do not require ontological correlates. It is possible to know that water freezes at a certain temperature because certain molecules have such and such properties. How could these explanations be formulated if reference to concrete properties is not allowed? To use Armstrong’s example: let us suppose that a kettle of water is heated. What causes the heating of the water? According to Armstrong, the fire must be in the right relation to the kettle, say underneath it, and the kettle must contain water. The fire must be hot in order to cause the heating etc. As Armstrong notes:

Consider how this is explained by an account in terms of predicates. The predicate ‘underneath’ applies to the pair of the fire and the kettle, the predicate ‘hot’ to the fire and, eventually, to the water. But when we have said that these predicates apply, we have surely not said enough.

A nominalist has to explain why it is the case that a certain predicate “fits” on a given occasion. Both Davidson and Quine have an answer. Quine’s answer is: “[…] that houses and roses and sunsets are all of them red may be taken as ultimate and irreducible”. This is not a clarifying answer. Davidson says: “If an evaluative word fits, it’s because of properties that we could describe in non-evaluative terms.” But how does this answer go together with the view that properties are abstract? If the term “property” is replaced with the term “predicate”,

---

563 Davidson, 1993b, 608.
564 Armstrong, 1992, 164.
566 Ibid.
567 Quine, 1960, 81.
568 In Bergström and Föllesdal, 1993, 218, my emphasis. On the other hand, Davidson (1974, 194) has famously claimed that: “Nothing… no thing, makes sentences and theories true….” This suggests that there need not be anything about events that explains why certain predicates apply to them.
then the answer is that an evaluative word fits because a non-evaluative predicate fits. How would *this* clarify why the evaluative word fits? The explanation would concern linguistic conventions and it would remain utterly obscure why the attribution of an evaluative term, say “good”, would have anything to do with the fact that a non-evaluative term happens to apply to the same object as the evaluative term.\(^\text{569}\) It is thus understandable that Davidson gives the explanation in terms of *properties*.

But why use the term “property” if everything could be explained without it? It is not easy to see how Quine and Davidson can deny (of course they do) that objects satisfy certain predicates because of such and such physical facts (which are in the object); but if they are not denying *this*, then what are they denying? Recall the example of the *microscopic properties* of the chalk. Doesn’t this formulation imply that the chalk breaks because of the properties it possesses? Being a grandfather, for example, can be thought of as a relational instead of a microscopic property of a man, but how could it be denied that the *microscopic properties* of the chalk are not properties *intrinsic* to it? If these properties are intrinsic, then they are in the chalk and the predicate of, say frangibility, applies to the piece of chalk *because* of them. Davidson has claimed that a fully adequate science “[…] relies on knowledge of the structure that explains why [a] brittle object shatters when it does, or allows us to predict when, and to what degree, an object will return to its original shape after being subjected to precisely specified forces”.\(^\text{570}\) How else could the structure be understood than as a concrete physical property? The following statement of Davidson’s surely gives the impression that properties are, after all, something concrete:

> […] if an object has a disposition, this fact must depend on the physical properties of the object. So whatever can be explained by appeal to the disposition must be explicable in physical terms. Solubility illustrates the point: at one time we knew there was some unknown physical property of an object that made it soluble; now we know what the property is.\(^\text{571}\)

Davidson is here describing Quine’s view, but given the nominalism of the latter it cannot be the case that Davidson, who shies away from nominalism, would be *less* realistic about

\(^\text{569}\) The problem is obvious in the mind–body case. Psychological terms fit because of properties (or predicates) that could be described in non-psychological terms. But if the explanation stops here, the connection between consciousness and brain-states remains a mystery. Whereas the view that the applicability of the predicate “good” or “beautiful” cannot be explicated in non-evaluative terms may sound plausible, the view according to which the physical property which corresponds to psychological predicates could not be explicated sounds implausible. Whereas there may be nothing in the object which makes us to judge it as being beautiful, mental properties are in the subject because of physical properties. Of course, all this is true only as long as one believes in the truth of O-physicalism. I am not defending O-physicalism here but merely discussing how things *should* look if this version of physicalism is accepted.

\(^\text{570}\) Davidson, 1993f, 311.

\(^\text{571}\) Davidson, 1995a, 120.
properties than Quine. I therefore conclude that the above describes also Davidson’s own view.\footnote{Indeed, similar kinds of statements are given in Davidson 1991 where Davidson certainly describes his own position.} Obscurities follow if in the above quote the term “property” is changed to the term “predicate”. What would it mean to say that there was some unknown physical predicate that made an object soluble and now we know what the predicate is?

It is problematic that Davidson uses the terms “predicate” and “property” interchangeably. This suggests that he accepts properties in his ontology. Usually Davidson talks about properties instead of predicates, and this is an important reason why so many commentators continue to discuss his position from the perspective of property realism. For example, in the quote above the claim that disposition must depend on the physical properties of the object easily leads the reader to assume that, according to Davidson, there are such things as properties. How could something depend on physical properties if these properties are abstract and nowhere? But, of course, if Davidson denies the concrete existence of properties, then it is no wonder that he claims that mental properties do not reduce to physical properties – because what he is in fact saying is that mental predicates cannot be reduced to physical predicates. This in turn can be understood as a claim that mental predicates cannot be defined in terms of physical predicates, and this is one central claim of AM. Mental predicates do not translate into physical predicates. If this is the essential lesson, then the label “property dualism” is inappropriate and Davidson’s position is best understood as a form of predicate dualism. Then the “dualism” of Davidson’s position means “merely” that mental talk cannot be translated into talk which uses only physical terms and that mental talk is not an optional part of our conceptual resources.

However, let us for the moment proceed under the assumption that Davidson is an antirealist with respect to properties. He notes: “[…] properties are abstract objects. They are nowhere.”\footnote{Davidson, 2000a, 24.} Since Davidson does elaborate his claim, let us not speculate about the question of what abstract objects are, but instead take Davidson’s claim that “properties are nowhere” as conclusive evidence for the conclusion that, in his opinion, “properties are not anywhere in any clear sense”. But what are the reasons to say that properties are, pace Armstrong for example, nowhere? Although Davidson is not a nominalist, his commitment to Tarski’s theory of truth has the consequence that Davidson has no use for properties.\footnote{For a discussion of this aspect of Davidson’s view see especially Neale, 2003.} He notes:
I am firmly convinced that when we make such remarks as ‘Alphonse believes that Lisbon is in Portugal’ what we say about Alphonse is objectively true or false. Statements like this attribute a genuine psychological property to a person. People believe, want etc. this and that. So far everything is understandable. Statements about psychological states are objectively true or false just like statements about non-psychological states, because the former attribute a genuine psychological property to a person. We can notice that on Davidson’s view mental properties would be properties of a person, not properties of an event (as Kim thinks) or properties of the brain (as Churchland thinks). But then the story is spoiled.

Actually I don’t want to invoke properties at this point either: attributing a property is done by using a predicate, and the semantics of predicates don’t need or want properties. Having banished properties, states, beliefs, and so on, from my explanatory machinery, I will now continue to allow myself to use words that superficially seem to refer to such things, on condition that I be understood to hold that such talk can be exchanged at boring length for talk that doesn’t use these words as referring to entities. It should be clear that my ontological scruples are due only to my suspicion that these entities do not help explain what I want to explain.

So, in the end the reason why Davidson does not want to talk about properties is his view that everything that could be explained by using properties can be explained by using predicates. This view obviously faces the difficult challenge of whether all talk and every explanation that refers to properties can be exchanged with talk that does not refer to them. This problem cannot be considered here; yet the problem raised by Armstrong is something for which a nominalist would be required to give an answer. I believe we can discuss Davidson’s view about properties without trying to solve, beyond what we have already said, the question of whether properties exist or what our attitude towards properties should be.

We now know the reason behind Davidson’s antirealism about properties. There is no clear semantic need to treat states or properties as entities, that is: “[...] the semantics of predicates don’t need or want properties”. Indeed, as Davidson says “there are no compelling reasons to introduce entities corresponding to predicates in giving the semantics of predicates”. Davidson’s skepticism towards properties is a result of a project in the philosophy of language where he tries to provide an adequate compositional meaning theory for natural languages. This project can, according to Davidson, be accomplished without assuming entities which correspond to predicates. As the previous quote shows, the semantics of predicates do not require entities corresponding to predicates. The facts about the use of predicates can be explained without introducing entities corresponding to these predicates. In

---

575 Davidson, 2001d, 298-299, first emphasis mine.
576 Ibid. My emphasis.
577 Davidson, 1993g, 288.
their recent study of Davidson’s philosophy of language, Lepore and Ludwig point out precisely this: “No properties… of any kind are required to carry out the project of providing an adequate compositional meaning theory for natural languages for sentences which do not explicitly refer to them.” Many commentators see Davidson’s commitment to Tarski’s theory as the key to his rejection of properties. The questions that were asked in the beginning of this section, namely what Davidson’s attitude towards properties is and what the reasons for it are, can now be answered. Davidson does not want to invoke properties because of his project in semantics. This much is clear.

Commentators have nevertheless drawn very different conclusions from Davidson’s Tarskian and Quinean commitments. Some, like Campbell, Neale, Lepore and Ludwig claim that it is precisely Davidson’s commitment to Tarski-style semantics which leads him to say that there are no properties in virtue of which events support certain descriptions. Others like Crane and Marras claim, without going into details, that the issue is broadly one of “Davidson’s nominalism”. Their conclusions about what Davidson would say about the reality of properties vary. According to McLaughlin, “it is far from certain whether… Davidson thinks there are any properties at all, mental or physical”. Crane is more explicit: “[…] on Davidson’s… nominalistic view there are no properties”. A recent conclusion by Gibb is that for Davidson, “properties are not objective aspects of things in the world”. Melchert concludes that according to Davidson there are no mental properties. Campbell seems to agree with this, since he notes that Davidson is “perhaps a realist” with respect to physical properties but an anti-realist with respect to mental properties. Yet, in a similar vein as McLaughlin, he notes: “Whether or not one can show that Davidson has implicit realist commitments to properties is unclear.” This view suggests that Davidson’s view about properties is something which cannot be settled with certainty. All the commentators cannot be right, but who is?

578 Lepore and Ludwig, 2005, 8.
580 Crane, 1992, 2000, Marras, 1999. As I noted already, this interpretation is, to say the least, interesting given Davidson’s own insistence that he was not a nominalist. The interpretation should thus be treated with caution. Commentators seem to disagree also about the extent of Davidson’s nominalism. Evnine (1991) notes that there is a “nominalist streak” in Davidson’s work, whereas Glock (2003) claims that Davidson does not subscribe to nominalism.
582 Crane, 1992, 189.
584 Melchert, 1986.
2.3.4 Is Davidson’s rejection of properties justified?

In the previous section I described why Davidson is “suspicious” of properties. Now we turn to the question of whether Davidson’s commitment to a Tarski-style semantics gives a good reason to claim that properties are unreal or that they are nowhere. Some of the problems that are currently being discussed in the philosophy of mind, like the problem of mental causation, may require a discussion of properties. It is not obvious that from the fact that one does not “need” properties in his philosophy of language, it would follow that he also does not need properties in his philosophy of mind. Tarski, to whom many refer as the source of “Davidson’s nominalism”, notes:

[…] we may accept the semantic conception of truth without giving up any epistemological attitude we may have had; we may remain naïve realists, critical realists, or idealists, empiricists, or metaphysicians—whatever we were before. The semantic conception is completely neutral towards all these issues.

Davidson himself admits that deep metaphysical problems about events remain and that the study of the logical form of sentences “can carry us only a certain distance.” Robustly ontological questions about the status of mental properties remain. In fact, as Davidson notes, there are two separate questions about properties:

The first is whether there is in general any objection to including properties in our ontology. The second is whether there is any advantage in introducing predicates as the sole semantic items used to explain the function of predicates.

It is the second question which captures Davidson’s main interest, since with respect to the first question he notes: “I have no objection to accepting properties, if they turn out to be needed to explain such sentences as ‘This is the same colour as that’”. Why should not this attitude be applied also to mental properties? Should we not accept such properties if they turn out to be needed to explain sentences such as “Donald has the same belief as Georg”? There is

---

586 Davidson obviously thinks that these two branches of philosophy cannot be clearly separated. There is no reason why one would have to accept this view as strongly as Davidson does. In fact, the problems that discussed here suggest that a satisfying answer cannot be found if one models one’s views about the mind solely in terms of the considerations that he puts forward in the philosophy of language. The questions of “What are mental properties?” or “How is mental causation possible?” are robustly ontological. This is, at least, the common opinion in the contemporary philosophy of mind.

587 Tarski, 1944, 362.
588 Davidson, 1980a, 146.
589 Davidson, 1999r, 88.
590 Ibid.
no reason to be skeptical about such properties in these kinds of cases. In his last statement on the subject Davidson notes:

Tarski’s truth definitions, and theories of truth based on them... distinguish between the issue of whether properties and other abstract entities exist and the semantic role of predicates. Of course, if some (second-level) predicates are true of abstract objects, those objects must exist...

Given Davidson’s view it follows that if certain sentences cannot be dealt with without positing properties, he would allow them into his ontology as long as these properties are treated as *abstract* objects. Commentators have taken this into account. Crane, for example, notes that nominalism *a la* Quine and Davidson does not need to deny the existence of *all* abstract objects. Davidson himself affirms that “no one has good idea of how to do syntax or semantics without appeal to [abstract objects]”.

Since Davidson accepts properties in the sense described above, the real question is what kind of “things” these properties are. I will not go too deeply into metaphysical questions, but nevertheless but the following should be recognized. According to Rosen: “[...] if any characterization of the abstract deserves to be regarded as the standard one, it is this: An abstract entity is a non-spatial (or non-spatiotemporal) causally inert thing.” It is impossible to tell whether Davidson’s use of the term “abstract object” fits this characterization. However, Davidson does claim that properties “have no location” and that “they are nowhere”. The non-spatiality of abstract objects seems thus to be satisfied. What about the causal inertness? This would seem to follow from the non-spatiality of properties. How could something which is nowhere have any causal effects on anything? Is it even appropriate to use the term *something* in the case of properties if they are abstract and not anywhere? Armstrong and Heathcote argue that: “Davidson’s Quinean nominalism does not enable him to say that it is the properties of cause and effect are the main players in causation.”

---

591 But it is crucially important to understand correctly the status of mental properties. That people share a mental property means that their states are sufficiently similar for us to understand them, not that they share an underlying physical property.

592 Davidson, 2005a, 158.


594 Davidson, 1999s, 101.

595 Rosen, 2001. The reference is to a web-page.

596 Let us suppose that according to Davidson mental properties are abstract objects and thus nowhere. Von Wright makes an interestingly similar claim when he notes that: “Meanings are no-where. So, if the mental were the meanings (of behavioural signs caused by nervous processes), then mental things would not exist in space, i.e. physical space” (Von Wright, 1997, 133). There is a hint of intentional antirealism in the views of both Davidson and von Wright. I will consider this aspect of their views in section 2.5.3.

597 Armstrong and Heathcote, 1991, 68.
Davidson cannot say that properties are literally a cause. Crane’s reading will be questioned in section 4.1, but let us note here that according to this “nominalistic interpretation”, properties are not causally efficacious. In fact, Davidson himself says that properties are causally inert and he goes on to say that it makes no sense to say that A causes B in virtue of having such and such properties, and finally he claims that abstractions generally have no causal relations. But what is the point of positing the existence of abstract entities then? Many philosophers think that if mental properties are abstract then they cannot be causally relevant, and this being the case it is not clear why it would be worth having them at all. Crane has noted: “[…] properties understood [as abstract entities] cannot be causes, since causes must have spatial or (at least) temporal location”. It is interesting to consider whether von Wright’s temporal supervenience would grant mental phenomena a causal role, for example, on Crane’s account. This would be another interesting direction in which von Wright’s position could be developed: by its insistence on temporal supervenience, this position could be taken to provide a positive solution to the problem of mental causation. We should note that if we take Davidson’s line of thought, then properties are nowhere and they do not seem to have causal powers, and their causal inertness leads to their elimination. There have been many accusations that Davidson’s theory leads to epiphenomenalism. I will consider these charges in chapter four, but it should be noted here that this question is closely connected to Davidson’s views about properties. The problem of mental properties, so it is claimed especially by Kim, is a problem for non-reductive physicalism because of the problem of mental causation.

What can thus be said about Davidson’s rejection of properties? Although it seems that Davidson is an antirealist about properties he, quite surprisingly, remains silent when it comes to finally saying something specific about their ontological status. Are there such things or not? Davidson comments: “Properties are abstract, that is, if they exist. About this I am agnostic.” If Davidson is agnostic about the question of whether properties exist, why have so many commentators concluded that he is an antirealist about properties? This cannot be anything else than a misreading. Davidson notes: “There is no objection to taking properties and relations as entities about which we want to think and say things, unless, of

[600] For this worry see, for example, Dretske, 1989.
[602] Davidson, 2000a, 24, fn. 7. According to Davidson, predicates are abstract objects too.
course, there are no such entities. I shall not cast doubt on their existence.”603 In his reply to Jennifer Hornsby, Davidson notes that some may think that his reason for talking in terms of predicates instead of properties might be the desire to avoid references to universals.604 But, as Davidson notes: “I have no objection to referring to universals when it promotes some explanatory or semantic project…”605 These claims certainly create doubts about all those interpretations which simply declare that Davidson rejects properties altogether. Yet it must be noted that Davidson nevertheless firmly thinks that if properties exist they must still be abstract. In my opinion it could be said that Davidson takes an “easy way out”, because when it comes to properties:

[...]I plan to concentrate on what might be called the epistemological problem and let the ontological problem, if there is one, take care of itself. For I think that if we were to solve the epistemological problem we would lose interest in the supposed ontological problem.606

Davidson is willing to admit properties into his ontology because they “take up no space”. This is an interesting suggestion and an example of the way in which Davidson thinks about the relationship between epistemological and ontological questions. According to Stephen Neale, “Davidson suggests that if the need to posit a particular ontological category does not arise in the construction [of an acceptable T-theory], then the need cannot arise at all.”607 This interpretation emphasizes the relevance of the analysis of language for ontological questions. But what does it mean to say that there is no need for a particular ontological category? How to decide whether there is a need or not? If properties are needed for certain explanations, then it would seem that there is a need for properties. Whether they are needed is a question over which “intuitions” can differ. One conclusion that can be drawn from Davidson’s “agnosticism” about properties is that he is not an antirealist. Another conclusion is that it is not clear after all what Davidson’s stance with respect to the question of the reality of properties is. It is therefore not surprising that critics have held different opinions about this matter. The fact that Davidson does not state clearly his views about properties is an important observation, since critics have noted that Davidson’s view about properties is, depending on the critic, either a solution to the problem of mental causation or gives rise to this very problem.

603 Davidson, 2005, 84.
604 Davidson, 1999p.
605 Davidson, 1999p, 636-637.
606 Davidson, 1995b, 44.
What I have shown is that Davidson’s view about properties cannot simply be interpreted as being antirealist or realist. This is to be expected, since his version of intentional realism does not fit the standard realism / antirealism distinction well either. This being the case, it is not surprising that his view on the status of mental properties is likewise not easily fitted into the realism / antirealism dichotomy either. Perhaps Davidson’s position is best described as epiphrastic, as one commentator, Welshon, has done. According to the epiphrastic view of properties, properties are not to be distinguished from concepts or predicates. It is quite clear that Davidson subscribes to this view, but it is not clear how the view should be understood. Armstrong has noted:

To appreciate the utter implausibility of the attempt to evade properties by means of predicates it is perhaps sufficient to consider a case where a thing’s property changes. A cold thing becomes hot. For one who puts his or her faith in predicates this is a matter first of the predicate ‘cold’, or its semantic equivalent, applying to or being true of the object, and, second, the predicate ‘hot’ becoming applicable after cold loses applicability. Properties in the object are but metaphysical shadows cast on that object by the predicates. But what have predicates to do with the temperature of the object? The change in the object could have occurred even if the predicates had never existed.

Armstrong’s conclusion is: “[…] one has to be pretty far gone in what might be called linguistic idealism to find predicates much of substitute for properties”. I think we should conclude that Davidson does not offer good reasons for thinking that there are no such things as properties. His general rejection of properties (if that is what he is doing) is not sensitive to the kinds of questions which can be taken, with some justice, to be relevant in the philosophy of mind. Examples of the issues which are relevant but difficult to answer if an epiphrastic view about properties is accepted are, to mention only few:

1) If the relation between a mental event and a physical event is the relation of identity, it is still possible to ask why a certain physical event is also a mental event. There must be an ontological answer to this question as long as one does not accept the view that events are “mental only as described” in the strong sense, i.e. in the sense in which a distant collision of stars counts as a mental event. A very general way in which this question can be formulated in terms of properties instead of events is to say that consciousness is a mental properties.

---

608 Welshon, 1999.
609 Davidson’s acceptance of this obscure view has the unfortunate consequence that certain commentators keep interpreting his views without problematizing the epiphrastic view. For example, Simone Evnine (1991, 68) in his book about Davidson writes: “In what follows, I shall take supervenience indifferently to be about mental predicates or mental properties.”
611 Ibid.
property. But why are certain brains conscious? It is not a satisfying answer to say that brains are conscious only as described; there is (and could we say must be) a difference between a conscious and a non-conscious brain, a difference which can be found from the brain. This is a simple matter of scientific truth.

2) If mental events really cause something, we may want to have an ontological answer as to why this is the case. “Ontological answer” here means an answer which is formulated in terms of causally efficacious physical properties.

3) To what extent can the study of the brain tell us something about the mind?

4) Where are mental events if their existence is just a matter of a mental predicate fitting?

And so on. I am not saying that these are the essential questions that one should be asking in the philosophy of mind. I am not even saying that I take these questions to require answers which depend on the acceptance of properties; quite on the contrary. I merely note that many philosophers would raise these kinds of questions against an epiphrastic view. I believe that a Davidsonian and von Wrightian way to think about the mind suggests that some of these questions may be ill-formed. Yet, at least some of these questions sound intelligible and a dismissal of them because of obscure reasons given for antirealism about properties is perhaps a rushed dismissal.

Does the following view of Davidson offer a satisfying description of the nature of mental properties? Consider:

If you ask what kind of properties we’re attributing when we attribute beliefs, I think a theory about how we tell that belief-attributing sentences are true provides the best answer. This shows what kind of property it is: it’s a property which you determine to apply to an individual in the following way... (and here you describe the method). Is there something more to say about it? I don’t see why there has to be.

---

612 I believe we can understand this claim without taking a stance on the question of the nature of properties generally, or of the status of mental properties particularly.

613 An important question is how much emphasis is put on the scientific truth. What is however clear is that both Davidson and von Wright respect these truths, as their naturalism requires. This being the case they would accept the view that the reason that the property of being conscious is exemplified by a person would depend on the brain of the conscious person.

614 Davidson, 1993c, 196.
This view about mental properties reflects the Wittgensteinian view, which was briefly described in section 2.3.1. What needs to be considered is how we come to apply mental predicates to others; this captures something essential about the nature of mental properties.

I believe it possible to think that there could be something “more to say” about mental properties than the description of a theory which tells when mental predicates are true of a subject. It is not clear how Davidson’s remark that “there does not have to be anything more to say” about properties should be understood. Either there is something more to say or there isn’t. Many contemporary philosophers, for example Kim, would claim that the view given in the quote above neglects important questions. The question about the relation between mental and physical properties is precisely one of these questions. What, for example, can be said about the reality of mental properties if properties are abstract objects that have no location? Moreover, we should keep in mind that Davidson uses (all the time) expressions like physical (or mental) “features”, “respects”, “characteristics” and, of course, very often the term “property” itself. It is unclear how these terms should be understood if the only sense of property-talk is the one described in the previous quote. It seems to me that Davidson accepts property-talk when it suits his purposes. This is most evident in his discussion about the possibility of mental causation when he talks about properties without quarrels. As I showed in section 2.3.3, Davidson uses the expression “physical property” without hesitation and in a way which suggests that he thinks of physical properties as something concrete. The contexts in which such expressions are used do not allow for an epiphrastic understanding of properties.

If properties are treated as abstractions, or if it is claimed that properties should not be distinguished from concepts, attempts to answer certain questions in the philosophy of mind face serious problems. This of course is a satisfying conclusion for some. The conclusion is that the relation (or non-relation) between mental and physical properties cannot be further explicated and therefore a reductive analysis of mentality is an impossibility. But perhaps one, in an anti-Wittgensteinian “metaphysical mood”, feels that there has to be something more to say about the nature of mental properties than what Davidson admits. Perhaps the view that there has to be something more to say is a result of an uncritical acceptance of the metaphysical framework of physicalism. The temptations of this framework need to be

---

615 See Kim, 2003.
616 Davidson’s view about the attribution of mental properties does not clarify the question of whether these properties could be understood as physical properties. Whereas the view may suggest what kind of properties mental properties are, it remains silent on the question of whether they are physical properties and does not imply the non-physicality of mental properties.
considered carefully, but I also think that they cannot be completely ignored. Although I believe that a Wittgensteinian perspective should be more strongly used against physicalism, it seems to me that certain claims of O-physicalism sound plausible. As von Wright has claimed, certain views are natural and empirically plausible; for some the view that there must be something more to say about the nature of mental properties than what Davidson suggests may be very natural. In Davidson’s case it should also be acknowledged that, given his physicalism, it would be surprising if the nature of mental properties could not be explicated in any way which refers to a physicalistic ontology and uses the resources of this ontology. Despite Davidson’s Wittgensteinian conception of philosophy, he also accepts many basic principles of physicalism.

2.3.4 Are mental properties physical properties?

I have shown what Davidson’s view of properties is, and concluded that the reasons he gives for treating properties as abstract are not satisfying. The reasons he gives for rejecting properties once and for all should also be treated with suspicion. I nevertheless emphasize that these reasons may be satisfying if one accepts that semantic reasons to believe in the non-existence of properties override other reasons to believe in their existence. Davidson says that we would lose interest in ontological problems if epistemological problems were solved. But we have also seen that, according to him, deep metaphysical problems remain even after the logical form of sentences have been studied and clarified. The final answer is thus partly a matter of intuition.

To the question “Are mental properties physical properties?” Davidson would thus answer negatively, because properties are not part of his official ontology. Because the reasons for the rejection are unsatisfying, I believe we should consider more generally whether non-reductive physicalism can make sense of the view that mental properties are physical properties. A positive answer would remove one major problem haunting non-reductive physicalism – the charge of property dualism. An answer to the question about the relationship between mental and physical properties tells something important about the mind. All this being said, I note that Davidson’s view on properties is not entirely clear. As I showed in section 2.3.3, commentators have drawn different conclusions from the same textual evidence. Most critics see Davidson’s rejection of properties as being a result of his nominalism; yet Davidson claims that he is not a nominalist. Moreover, commentators have differing opinions about the properties that Davidson is allegedly antirealist about. Some
argue that he is a realist about physical properties but antirealist about mental properties, while others claim that he is antirealist about mental and physical properties.

These different interpretations of Davidson’s position have created confusion in the subsequent discussion concerning the questions of whether AM can account for mental causation or whether AM is a version of epiphenomenalism (or leads to it). Kim notes that Davidson is not comfortable with unrestrained talk of properties, and although Kim emphasizes that Davidson-interpreters may need to be careful with respect to this, he nevertheless notes that he will continue to discuss Davidson’s position in terms of properties because “property talk has been the norm”.617 This creates a problem which some commentators have recognized. If Davidson rejects properties, then an accusation of epiphenomenalism, which is grounded on the claim that the causal efficacy of mental properties has not been shown, is simply ill-founded. In other words, commentators criticize Davidson while accepting the background assumptions which Davidson himself would reject right from the beginning.618 This problem has been recently emphasized by Gibb.619 Criticisms based on assumptions which Davidson would reject result in an unfair description of Davidson’s position. I believe that an important part of the problem of property-dualism is a result of a similar kind of confusion. Certain philosophers, like Kim, have a view about mental properties, and what they are really doing is accusing non-reductive physicalists of the fact that they do not share this view.

As far as Davidson’s position is concerned I suggest that, given the textual evidence, Davidson is a realist about physical properties (physical predicates are ontologically grounded), and that if he is a realist about mental properties too, his only route to avoiding robust emergentism (which von Wright perhaps accepts) or full-blown property dualism (a la Chalmers) would be to say that mental properties are, after all, physical properties. Since Davidson is not willing to do this, I shall consider the question about the relation between mental and physical properties independently of his views. The conclusion will be that a non-reductive physicalist has no alternative than to admit that mental properties are physical properties. But there is no reason why a non-reductive physicalist could not admit this while still accepting the essential principles of his non-reductive position. We should disagree with

617 Kim, 2003, 135.
618 This is obvious, for example, in the debate between Davidson and some of his critics in Heil and Mele, 1993.
619 Gibb, 2006. Unfortunately Gibb herself makes the same mistake of which she accuses other commentators. After criticizing those who have a different view about properties from Davidson’s, she herself argues against Davidson’s view based on her own ontological reading of properties.
Davidson’s view that solving the epistemological problem of properties (the problem of *predicates*) clears all ontological worries about properties. The problem of properties certainly *seems*, above all else, to be an ontological problem. Whether there are such things as properties and what their nature is, is a question about the nature of reality. This is how the problem is usually treated in the contemporary philosophy of mind. We attribute mental properties to people and are interested, for example, in the question of what the relation between these properties and the robustly physical properties of the brain is. This is an ontological question and a claim according to which *this* question can be completely dismissed in favor of the epistemological problem, which is the problem of how we attribute mental properties and how to tell whether the attributions are true, is not a satisfactory answer to the ontological question. Gibb has explicitly raised this objection against Davidson’s position. She writes:

[…] from an ontological point of view, Davidson has arguably got things the wrong way around. It is arguable that one’s motivation for accepting or rejecting an ontological category should not have semantic considerations at its base, because contrary to Davidson, a theory of meaning cannot be appealed to in order to settle ontological issues. An appropriate semantics is to be construed only after the ontological issues are settled. Given this approach to metaphysics, whether or not one should admit the ontological category of properties is not to be based upon consideration of whether sentences require quantification over properties…rather, whether one should admit the ontological category of properties, and indeed what properties are, is to be established through metaphysical enquiry…

There are several very interesting things worth considering in Gibb’s suggestion, which is a good example of certain kind of approach that has been taken towards the problem of mental causation in the contemporary philosophy of mind. This approach could be labeled “overtly metaphysical”. Gibb, Lowe and others have a research agenda which is based on the view that the mental causation debate has been “framed with insufficient metaphysical precision” and that the purpose of their research is “to explore recent advances in metaphysics, in particular new accounts of the categories of being and of levels of being.” It is telling that Gibb uses the term “from an ontological point of view”, which could be taken to be the opposite perspective from Quine’s *logical point of view*. It is unclear what the ontological

---

621 See, [http://www.dur.ac.uk/philosophy/ontologyofmentalcausation/](http://www.dur.ac.uk/philosophy/ontologyofmentalcausation/). The group believes that metaphysical enquiry would constitute a significant step towards the resolution of the problem of mental causation. Is this project lost in the jungle of metaphysics of which Wittgenstein and von Wright warned? It is not a scientistic jungle of metaphysics, but it (perhaps) is a jungle in which the philosopher is no longer *looking* at the way concepts work in our life and thought, but is constructing detached views about how they *must* work, given certain structural assumptions.
622 Quine, 1953a.
point of view is exactly, but the view can be achieved through “metaphysical inquiry”, which “settles ontological issues”.

Whether or not Gibb’s view is reasonable, it is nevertheless a very unfair treatment of Davidson’s position. To claim that “a theory of meaning cannot be appealed to in order to settle ontological issues” or that “metaphysical inquiry should settle whether an ontological category should be admitted” is merely to say that this is one way to think about the issue and that Davidson’s view is incorrect because it differs from this specific way. To say that ontological issues can be settled through metaphysical inquiry is the opposite of what Davidson claims. Gibb does not elaborate why a theory of meaning could not settle ontological issues, and she does not explain how a “metaphysical inquiry” should proceed.

Although the kind of criticism raised here is unfair insofar as the critic does not explain why a theory of meaning could not settle ontological issues, it is at the same time clear that Davidson’s analysis of mental properties is not an answer to the question of whether mental properties are the properties of the brain – which in the end is the way how the problem of mental properties is usually formulated. The substantive question in the contemporary philosophy of mind is that of whether consciousness is reducible to neural processes, and Davidson’s view cannot help to answer this question. Of course, whether this should be the substantive question is one of the most pressing issues defining the gulf between non-reductive and reductive physicalists. Davidson and von Wright may be correct in their view that the question which Kim proposes is not philosophically substantive. On the other hand, if one thinks that the question about the actual reducibility of consciousness is philosophically interesting then one is apt to be dissatisfied with Davidson’s tendency to ignore the metaphysical problem.

If Davidson’s view is seen as an answer to the question of what the relation between the mental properties and the physical properties of the brain is, then the relevant consequence is negative: mental properties are not properties of the brain. In the context of O-physicalism, this sounds counter-intuitive. Twisting Davidson’s words a little, he can be interpreted as saying that when we solve the epistemological problems about the mind, we will lose interest.

Incidentally, From an Ontological Point of View is the title of Heil, 2003, which considers the fundamental questions of what there is. Again, this way to consider what there is differs crucially from Quine’s view, shared by Davidson, of how to settle the question of “What there is”.

A somewhat similar kind of unfair treatment can be found in Antony, 1989, in which it is surprisingly claimed that the view which Davidson should hold is the kind of view suggested by Fodor.

The explanation of metaphysical inquiry is vague: “[a] consideration of whether [ontological categories] could exist (and thus their existence and identity conditions), whether they play a nonredundant role in one’s ontological system, and whether they can co-exist with the other ontological categories that one includes within this ontological system.” (Gibb, 2006, 420.)
in the supposed ontological problems about the mind as well. If the problems concerning
the ways that mental predicates are attributed are solved, there do not remain any interesting
ontological problems any longer. Given Davidson’s understanding of ontological and
conceptual problems and the connection between them, this sounds like an acceptable
strategy. However, it is not clear how well the strategy works when it is ontological problems
that one wants to study. Why believe that the interest in “supposed” ontological problems will
be lost? What does it even mean to say that we would lose interest in the “supposed
ontological problem”? Does it mean that, in a Wittgensteinian spirit, by solving the
epistemological problem we would see the ontological problem as dispelled? Would we
realize that there was no ontological problem at all? A brief overview of the topics in
contemporary philosophy of mind shows that ontological problems are at the forefront. This
shows that many philosophers have not lost interest in them. It could be argued that those who
still have an interest in such ontological problem(s) have not realized the consequences of
Davidson’s views for the philosophy of mind. This is what has been said about Wittgenstein’s
philosophy: certain problems keep tormenting contemporary thinkers because the lessons of
Wittgenstein have not been taken seriously. We can recall Rorty’s comparison: “Many who
have no use for Wittgenstein have none for Davidson, and for the same reason: to adopt the
views of either would be to dissolve problems which they have spent the best years of their
lives trying to solve.”

Rorty’s observation notwithstanding, I believe we should tentatively join those who
think that there are some interesting questions about the mind which are related to a
physicalistic ontology, such as what kind of a brain a brain “with” a mind is, and what it is
about brains as physical things which makes them “brains with minds”. There are brains with
minds and brains without them, and this applies also to human cases also. We are willing to
say that a person in a vegetative state is “mindless” although she has a brain. As far as we
know, mental events are lacking from the head of a person who is (properly) in a vegetative
state. Although our main evidence for this would be the behavior of the person in question, it
is plausible to think that the physical condition of her brain has something to do with the fact
that mental events have disappeared. If we suppose that O-physicalism is true, we should in

626 Perhaps this really is what Davidson thinks in a quite strong sense. As he (1999, 654) says: “Beliefs, desires
and intentions belong to no ontology…. “ If this view is taken literally, there is no ontological problem about
beliefs, desires etc. But does a thing which belongs to no ontology exist? Isn’t something which belongs to no
ontology a candidate for elimination?
627 Rorty, 2005. Reference is to an electronic review for which page numbers cannot be given.
628 Very young children and babies, certain mentally handicapped persons, certain psychotic people and some
Alzheimer’s patients (among others) would be borderline cases. They all have brains but who has a mind?
fact say that the condition of her brain has *everything* to do with the fact that mental phenomena have disappeared from her brain, and from this perspective the state of her brain would be “conclusive evidence” for the conclusion that her mental life has ceased. Isn’t the difference in mental capacities between a “normal” person and a person in a vegetative state an *ontological difference*, a difference which depends of the intrinsic physical properties of their respective brains? In the context of O-physicalism this claim sounds self-evident. As long as O-physicalism is accepted, we should have an interest in the ontological problem, or at least admit that it is still an unsolved part of the mind–body problem. On the one hand, I believe that a robustly metaphysical approach is mistaken and neglects Wittgenstein’s insights. On the other hand, it seems to me that a philosopher who accepts O-physicalism cannot brush aside all the problems which are related to a monistic, physicalistic ontology.

With respect to Davidson’s position, one should therefore agree with Kim when he argues that “it is important to appreciate some robustly metaphysical versions of Davidson’s many central doctrines.”629 One should agree with this view because there are interesting questions about the mind which are “metaphysical”, and one is interested in knowing how Davidson’s approach particularly, and non-reductive physicalism generally, faces them. The kinds of questions which one wants to ask if one is robustly metaphysical are dismissed if the discussion is taking part only at the level of predicates. Now, it is quite certain that Davidson would indeed have dismissed these kinds of ontological questions. This kind of attitude is a central lesson of AM. It says: Do not look at the mental through the lens of a physicalistic ontology, and do not think that you can, *without confusion*, ask in physical terms ontological questions about mental states. Events are mental *only as described*, so they do not have a “life of their own”, i.e. detached from the interpretative practices on the basis of which they are being attributed. Enough about ontology has been said when we have shown, through Davidson’s argument, that mental phenomena *are* token-identical with physical phenomena. If the argument structure which leads to token-identity is accepted, further *philosophical* discussion or *philosophical* speculation about the physical aspect of mental phenomena should stop. This I take to be the central point that von Wright is also trying to make.

---

629 Kim, 2003, 135. When discussing Gibb’s criticism I noted that this is a somewhat unfair reading of Davidson’s position. But my purpose is to find out whether Davidson’s view can cope with problems that many contemporary philosophers see essential and also in this sense to defend Davidson’s view. This being the case, Gibb’s claim that Davidson cannot be criticized from a perspective which he would reject at the outset, while somewhat correct, neglects the most important part of this kind of criticism. The issue is, of course, to find out whether Davidson’s position is tenable and the critics try to ascertain that by considering how the position copes with respect certain important problems of the mind.
This view goes, or can at least be interpreted as going, strongly against the idea that empirical research, the study of the brain, could tell us something fundamentally important about the connection between the mental and the physical. Although this is a philosophically interesting view, I believe that in order to remain interesting it should be strong enough to stand against the views of those who think that these questions are, by and large, empirical. Certain questions about the mind are metaphysical and trouble philosophers. I do not think that Davidson has completely succeeded in his attempt to just sideline these questions. One can try to ask interesting “ontological” questions about the mind while maintaining the central insights that Davidson offers about the nature of our mental vocabulary. However, it surely would be a merit of Davidson’s, as well as of von Wright’s philosophy, and an important consequence if it turned out to be the case that these ontological questions are indeed confused and should be dismissed. This possibility has to be kept open. I admit that in suggesting that one can try to ask interesting ontological questions about the mind, I may be falling precisely into the trap warned about by Wittgenstein: my suggestion may reflect the desire to ask and answer questions in the same way that science does. Von Wright warned of the “jungle of metaphysics” in which a philosopher may lose himself if he participates in a philosophic culture gone scientistic. But it is extremely difficult to reject all metaphysical questions about the mind. Why? Perhaps because we have a firm, although perhaps distorted, view of how philosophical problems should be solved. As I noted in section 1.3, both Davidson and von Wright participate in empirical speculations from time to time and – by doing this – are not always true to their conception of philosophy. This shows how difficult in the contemporary philosophy of mind it is to follow certain insights of Wittgenstein.

One interesting and obviously a robustly metaphysical question is: Are there mental properties? If the answer is affirmative, can these properties be understood as being physical properties? In section 2.3.1 I have suggested that mental properties are real, but it was also noted that a model of mental properties that is formulated by using a physical property as a paradigmatic example of a property is confusing. The assumed starting point was a common-sense understanding of mental properties, which shows itself at the level of ordinary talk. But are these properties physical properties? An affirmative answer to this question is not in any sense a result of common-sense reasoning, but is based on the acceptance of a scientific picture of the world according to which there is nothing else besides physical particulars and properties.

630 For Davidson our mental vocabulary is, of course, what captures the nature and complete essence of the mental.
But are we led astray by language when considering this question? This would be one way to see the debate that has emerged between reductive and non-reductive physicalists. Is the relation between mental and physical properties similar to the relation between higher level physical properties and lower level physical entities? Does it make sense to say that someone’s mental state of “being jealous” is constituted of physical entities like atoms arranged in a certain way in his brain? If not, the physicalistic understanding of the term “property” confuses us when the same term is used in discussions about mental phenomena. When it is said that a person has such and such mental properties, should we understand the term “property” differently than when it is said that a brick has the property of weighing one kilogram? Those who formulate the problem of mental properties as a problem of how these properties are related to physical properties may be wrong in their assumption that the term “property” can be understood in the same way when talking about mental and physical properties. Perhaps we can conclude that mental properties exist, but that they cannot be understood in terms of physical properties because they are two different kinds of “things”, even though the same term is applied to both. The reason why the same term is used in both cases is that there are similarities between mental and physical properties. This is, again, one of the lessons that can be learned from Wittgenstein.

If it is accepted that mental properties are physical properties, then it has to be the case that just as water’s liquidity is explained in terms of its molecular structure, a person’s being jealous could likewise be explained, for example, as his nervous system being in a certain physical state P and ultimately as being a certain formation of microscopic physical entities. Then we could use a microscope where this formation could be seen and assert: “This formation of subatomic particles is the jealousy of Donald.” I think we can at least vaguely understand what this statement means.\footnote{Putnam (1999) claims that this would not make sense. Malcolm shares this view.} A constellation of subatomic particles, a certain space-time location, would thus literally be a certain mental property. This constellation is, in turn, without doubt a physical property of an object. Recently neuroscientists have claimed that a certain parts of the brain correspond to the locations of certain mental functions; the claim that a specific part of the brain is a thought does not sound nonsensical to many scientists. We can claim that an object, Donald that did not exhibit a specific concrete constellation would lack one property that another object, Georg, that did exhibit it, would have. The constellation, being an objective physical property, would be something which could exist in different individuals at different times if the materials for the formation of the
constellation were present. If mental properties like “believing that water is wet” or “being in pain” are only a matter of an individual’s brain being in a certain state, then there is no principled reason why these properties could not be shared between different individuals. We can imagine that Donald thinks “water is wet”, and that this mental property of his is identical with a certain physical property P of his brain. Georg could have the very same property P, because the existence of P is just a matter of certain physical particles existing in a certain kind of combination in his brain. The existence of these combinations occurring in different places at different times is possible. We have seen that the argument for the multiple realizability of mental states or the argument for the anomalism of the mental do not show that this kind of view about the relationship between mental and physical properties is impossible or even implausible.

Is this merely a version of the type–type identity theory, or is the term “property” being used here in a looser way? If the latter is the case, then the claim seems to agree instead with the token-identity thesis; the specific belief “water is wet” at time t is identical with some physical phenomenon (be it a state, entity, or process). It is possible to say that at time t Donald has the property of believing “water is wet” that Georg lacks, and Donald has this property because at time t his body has a certain physical property P. But this way of using the term “property” is not the way that it is understood in the argument for type–type identity theory. Although this way of thinking about properties is not consistent with the sense of the type–type identity theory, it is nevertheless something stronger than the view that properties are epiphrastic. The reason for using the term “property” in a somewhat metaphysical sense (whether or not it here corresponds to the philosopher’s standard use of the term) is to question whether or not one should think of properties in an epiphrastic sense and think of properties as abstract objects. But can this talk about properties be understood also in the sense in which the term is understood in the argument for the type–type identity theory? It can, for it has not been shown that a mental property M could not be type-identical with a physical property P.

The problem for a Davidsonian is that since a type–type reduction is a priori ruled out, it can never be shown that a mental property M is identical with a physical property P. But the identity of M and P could be the case even if the possibility of establishing this is forever prevented. Epistemological irreducibility does not prevent the possibility of at least theoretical ontological reducibility, although it may prevent the possibility of establishing ontological identity. If actual type–type reducibility is impossible to achieve, then there is no rationale for insisting that a mental property M is identical with a physical property P; yet it is
still possible to claim that the property M is identical with some physical property P₁…Pₙ. To me, this seems to be something which an O-physicalist has to accept.

Nick Zangwill has proposed a view which could look promising for a non-reductive physicalist. He claims that the position is something that a Davidsonian could accept. Zangwill writes:

The real Davidson identified mental and physical events. But a Davidsonian might talk in terms of mental and physical states of affairs. A state of affairs is a structured entity consisting of an object or event which possesses a property at a time. Such states of affairs are particulars. We might then identify someone’s having a certain mental property at a time with his having a certain physical property at that time, as one might identify a mechanism’s being unlocked at a time with all its slots being lined up at that time. Or some might prefer it if we say that someone’s having a certain mental property at a time is constituted by his having a certain physical property at that time, much as a mechanism’s being unlocked at a time is constituted by all its slots being lined up at that time.⁶³²

Although it is unclear what “a Davidsonian” can tolerate in his metaphysics, it can be argued that replacing events with states of affairs, which are understood as particulars, does not affect Davidson’s arguments in favor of the irreducibility of the mental, which I see as the most interesting and important part of his philosophy of mind. From Davidson’s perspective the essential question is that of the usefulness of one sort of entity (e.g. states of affairs) over another sort of entity (e.g. events) in interpreting sentences that seem to be, for example about events. We must remember that Davidson’s “ontological scruples” are only a result of his suspicion that certain entities turn out to be “irrelevant” for explanatory purposes. But, semantic reasons aside, there seems to be no major difficulty for a Davidsonian in accepting the kind of view suggested by Zangwill. A Davidsonian could never show that Zangwill’s position is correct, but he could accept that this is how things nevertheless are.⁶³³ One could hold a view about the nature of reality while insisting that it is not possible to show this view to be true. A Davidsonian could accept the claim that mental properties are physical properties; yet he would have to insist that this cannot be shown. This would render the status of the claim that mental properties are physical properties quite empty, but it would be in harmony with the view that all phenomena are strictly determined and predictable. Every mental property would be identical with some physical property, never mind which one.

Because this work is not a study in metaphysics, the question about the actual physical mode of properties being mental must be left open. It seems that a notable part of the contemporary discussion concerning the problem of mental causation or the coherence of

---

⁶³² Zangwill, 1996, 72.
⁶³³ Such a Davidsonian would have to be somewhat more metaphysical or naturalistic than Davidson himself, perhaps carrying on the kind of metaphysical inquiry suggested, for example, by Gibb.

214
non-reductive physicalism ignores the deep metaphysical questions about the nature of properties. But, or so I believe, philosophers should make clear what they are saying when they talk about mental properties. They should engage more fully in the clarification of conceptual intuitions, to use von Wright’s expression. I would like to suggest very tentatively that together with Zangwill’s proposal, one could consider the trope-theoretic view of properties as a solution to the question of whether mental properties are physical properties, or better still, to the question of how best to understand the relation between these properties. Does a suggestion that mental properties are physical properties imply that the properties are identical? A claim that a mental property M is a physical property P seems to imply that they are in fact the same property. But does this make sense? Is Donald’s believing that “Water is wet” identical to the brain state in which he is at time t? This view would seem to face serious problems. But consider the suggestion by John Heil:

Soup is liquid in virtue of its possession of a certain molecular structure. We can also speak of the soup’s liquidity being realized in or by its molecular structure. The relation here is not a relation between types – properties or characteristics – but between instances or exemplifications of properties or characteristics, between the soup’s liquidity and a particular dynamic distribution of its molecular constituents. If this particular arrangement of molecules realizes the soup’s liquidity, then the soup’s being liquid is, I shall say, constituted by its molecular structure.

In this work I do not want to try to defend the kind of view that Heil proposes. I want to suggest, nevertheless, that one can think of the relation between mental and physical properties in terms of constitution or realization instead of identity. We do not have to consider the plausibility of this view, since for the purposes at hand it is enough to recognize that this kind of position is open to a non-reductive physicalist who accepts O-physicalism. As an ontological solution, a non-reductive physicalist could accept perhaps even the kind of realization physicalism proposed, for example, by Andrew Melnyk. At the same time, it needs to be recognized that the physical constitution or realization of mental phenomena must be understood somewhat metaphorically, as being simply a form of speech that relieves physicalistic anguish. It is difficult to make sense of a view according to which an individual thought (whatever that might be) is constituted of molecules.

I believe that a Davidsonian could be more robustly metaphysical than Davidson himself was. I think that Davidson himself could have accepted a more metaphysical view if it had not been for his other reasons for being suspicious about properties. A discussion about

---

634 These problems will be considered in section 2.4 and 2.4.3.
637 We come back to this problem in section 2.4.4.
the physical realization or constitution of mental properties would of course be overtly metaphysical from Davidson’s actual philosophical perspective. The kind of metaphysical inquiry in which, for example, trope-theorists are engaged seems utterly wrongheaded from a Wittgensteinian point of view. Whereas there are good reasons to share Wittgenstein’s skepticism towards this kind of enquiry, it should nevertheless be emphasized that a non-reductive physicalist with a non-Wittgensteinian conception of philosophy could explore different ontological options in order to answer the charge of incoherence put forward by reductive physicalists. It could be argued Davidson’s use of the concept of supervenience opens up the possibility for the interpretation that one way in which Davidson could have had discussed the relation between physical and mental properties is in terms of the former determining or constituting the latter. In section 4.1.1, Davidson’s understanding of supervenience will be considered in greater detail. Here I want to point out that statements like the following give the impression that the relation between mental and physical properties could be something along the lines that Zangwill and Heil, among others, have suggested.638

When Davidson first introduced the term supervenience he claimed:

Although the position I describe denies there are psychophysical laws, it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events alike in all physical respects but differing in some mental respect, or that an object cannot alter in some mental respect without altering in some physical respect.

639

He claims also:

Although… psychological characteristics cannot be reduced to the other, nevertheless they may be (and I think are) strongly dependent on them. Indeed, there is a sense in which physical characteristics of an event (or object or state) determine the psychological characteristics…

640

These are examples of passages where Davidson uses the term “characteristic” or “respect” unproblematically. It is also important to see that Davidson talks about the strong dependence of mental on the physical, and of the latter determining the former. I would say that dependence or determination, as notions, are as ontological as the constitution used by Zangwill and Heil. Although supervenience does not imply that the same physical properties change with the same mental properties, supervenience does not prevent this possibility either. Kim, for example, has claimed that supervenience is best understood as being a

638 For a discussion of determination and constitution of the mental by the physical see Polger, 2004 and Melnyk, 2003.
639 Davidson, 1970, 214, my emphasis.
640 Davidson, 1974b, 253.
reductive relation. When the concept was introduced by Davidson it was meant to describe a relation of dependency. Davidson writes, for example: “The property of being good is supervenient on physical properties in [the] sense that goodness wholly depends upon the physical properties.”641 How to understand this, given Davidson’s view of properties? It could be understood either as saying that the application of mental predicates depends on the physical properties of the object to which these predicates are attributed, or that the application of mental predicates depends on what physical predicates can be attributed to an object. According to the epiphrastic reading, Davidson would probably not make a distinction between these formulations. It is unfortunate that Davidson, because of his somewhat obscure reasons, did not discuss the different ways in which one could be a realist about properties. A trope-theoretic approach is perhaps something which Davidson could have had accepted. If we are more realistic about properties than Davidson, we can say that goodness, or any higher-level property, depends on the physical properties of an object in a sense somewhat similar to that in which the soup’s liquidity depends on its molecular structure, or in a sense somewhat similar to that in which an unlocked mechanism is constituted by all its slots being lined up. I say “somewhat similar” since there is no reason to believe that goodness could be defined in terms of physical properties or that one could nomologically relate goodness to physical properties. From this it does not follow that goodness is a non-physical property in any interesting sense. As Davidson himself notes: “The definition of supervenience implies that a change in mental properties is always accompanied by a change in physical properties....”642 It needs to be recognized that the talk about constitution cannot be taken literally, but I believe it is a good enough metaphor to make mental properties physically respectable.

What can be said of the position of von Wright? He did not really participate in the contemporary debate about the nature and status of mental properties and did not clearly express his general views about the nature of properties. Von Wright does note that the mistake of nominalism is to identify the meaning of a word with the reference of the word.643 In an unpublished manuscript von Wright nevertheless considered the question of how the claim about the identity of properties could be understood and made few interesting remarks.644 His general view was that the temptation to identify mental properties with

641 Davidson, 1964, 47, my emphasis.
642 Davidson, 1993d, 187.
643 Von Wright, 1998a.
644 A modification of the manuscript von Wright, UPg, was later published as von Wright, 1988. Interestingly the discussion in the manuscript about mental properties was left out from the published paper.
physical properties, or more broadly, mental phenomena with physical phenomena, must be
due to our inability to understand the ontological status of the mental in a correct way.
Considerations of this question lead us astray by presenting *misleading pictures*. Von Wright
thinks that is misleading or false to see consciousness as a *specific* phenomenon. Our
examples of mental phenomena are usually


This was stressed by Wittgenstein when he warned that the temptation to substantilize the
mind should be resisted. The tendency to model our understanding of mental properties in
terms of physical properties, or in terms which are first and foremost applicable to physical
phenomena, is still strong.

An identity claim often used as an example by philosophers is that of the identity of
“water” and “H$_2$O”. Another example used by von Wright is the identity of brittleness and a
certain microstructure. Von Wright claims that both examples of identities are problematic as
exemplars of the mental–physical relation. He notes that in the water-case two identity claims
can be distinguished. On the one hand, water is stuff which is a chemical compound of two
other stuffs. On the other hand, samples of water are aggregates of small bodies, molecules.
For this second identity claim, it is not essential that the molecules happen to be composed of
hydrogen and oxygen. Von Wright thinks that both identities are contingent. Water could
have been something else than H$_2$O, and therefore samples of water could have been
something else than compounds of hydrogen and oxygen molecules. This identity cannot be
used as an example of the mental–physical identity because there simply is no *something*
which could be both a brain state and a mental state in the sense in which a bucket of liquid
can be both water and a collection of H$_2$O molecules.

The assumed identity of brittleness and microstructure is problematic in a different
way. The property of being brittle is a different property than that of having of a certain
microstructure. They are two distinct properties of the same material object. However, there
has to be a substance of which both brittleness and microstructure can be properties. But,
again, in the mental–physical case, there is no substance besides the person which could be

---

645 Von Wright, UPf, 16, 22. See also von Wright, Upg.
the bearer of both mental and physical properties. The brain cannot be such a substance because one cannot attribute mental properties to the brain.

I conclude that an O-physicalist has no alternative than to claim that mental properties are physical properties. For the moment, notwithstanding the views of Davidson, we can conclude that mental properties are physical properties and the relation between them is a relation of constitution or realization. This is a solution which a non-reductive physicalist does not have to reject. Let it be emphasized, though, that I fail to see how the acceptance of the identity of properties would imply a stronger commitment to physicalism than the acceptance that every individual state is physical. What I fail to see, supposing that one has a proper understanding of the nature of mental properties, is how the question about the physicality of mental properties differs from the question of whether a specific state is physical or not. I fail to see this as long as it is supposed that no matter what our ontological views are, the irreducibility of the mental holds.

2.4 A problem for non-reductive physicalism: Token identity

It is unclear whether the problem of mental properties is a special problem for a non-reductive physicalist. What about the view which according to many captures the essence of non-reductive physicalism, namely the view that every particular is a physical particular. I think that the most serious problem that can be raised for non-reductive physicalism is: why believe this? If the claim about the physicality of particulars cannot be defended, then the position is not a very interesting form of physicalism; it is physicalism strictly by name only. But it seems to me that this is not a problem just for non-reductive physicalism. It is a general problem for physicalism, because whatever the view about properties is, it is clear that a physicalistic world-view cannot tolerate a world with non-physical particulars.646

The problem of whether token physicalism is true is a problem for non-reductivism and reductivism for different reasons. Let us suppose that it is being claimed that mental phenomena are physical phenomena and can be reduced to physical phenomena. Then two problems occur. First, we must show or at least suggest how the reduction could be done; otherwise the claim is empty. Second, we must refute the philosophical arguments which purport to show that such a reduction is impossible in principle. Of course, quite trivially, if it could be shown that, or more correctly how, mental phenomena reduce to physical

646 Of course “substantial” non-physical properties, for example properties occurring outside space-time, cannot be tolerated either.
phenomena, we could thereby largely defuse the philosophical criticism that this cannot be done. Since these kinds of reductions are non-existent at this time, the criticism that such reductions are impossible is relevant. Let us suppose, on the other hand, that it is being claimed that mental phenomena are physical phenomena but cannot be reduced to physical phenomena. Then we encounter a different problem. If the claim is that token-reductions are impossible, what is the reason to believe that mental phenomena are token-identical to physical phenomena? We must have an argument which is meant to demonstrate that mental phenomena are physical phenomena, although this can never be shown empirically. The argument cannot be empirical in any clear sense.

The purpose of this section is to consider what the reasons to hold the token identity thesis are, what the thesis actually means and, supposing that the thesis is true, how plausible actual token reductions are. These questions will be considered, as elsewhere, primarily from the perspectives of Davidson and von Wright.

### 2.4.1 A Davidsonian approach: Token identity between the mental and the physical

The essential claim of Davidson’s physicalism is that every particular is physical. But what does the physicality of particulars mean, and why believe it? Davidson gives one reason to back up this claim, which is the cornerstone of his physicalism. It is the principle of the nomological character of causality (PNCC), which states that causally interacting events must fall under a strict law. This, together with the other premises of AM gives, according to Davidson, a straightforward demonstration of the identity in question. How this is supposed to work was described in section 2.1.3. But in order to accept the demonstration, the truth of the premises of AM must be accepted. The thesis of the anomalism of the mental will be thoroughly discussed and defended in later sections. But what about the crucial premise for monism, namely PNCC? Yalowitz is correct when he claims that it is a key argument for Davidson’s monism; yet as was noted in section 2.1.3, Anscombe and Van Gulick, among others, raise the objection that Davidson has not given good reasons to believe the thesis. I have concluded that Davidson’s argument, which is meant to back up the principle, is obscure and unconvincing.

---

647 I will simply use ‘PNCC’ from now on.
Let us suppose for the sake of argument that one does not accept PNCC.\textsuperscript{648} This is an unfair interpretation of Davidson’s views, but let us accept it for the moment in order to test the coherence of a non-reductive view. What are the reasons to believe that every particular is physical if the principle is rejected? One would have a reason to believe this if one could show that each mental event is a physical event.\textsuperscript{649} If it could be shown, on each and every occasion, that such and such mental event at time $t$ is a certain physical event, there would be a reason to believe that mental events are physical events – at least until shown otherwise by a case, if such were to occur, where a correlation could not be shown. Token-identity would then be at least a well confirmed empirical hypothesis. A real correlation at time $t$ between a mental event $M$ and a physical event $P$ would not, however, show that mental event $M$ which would be “relevantly similar” to $M$ at time $t_1$ would correlate with $P$. The question about relevant similarity is obviously extremely complex: how to decide when temporally distinct mental events are relevantly similar? Some of these problems we encountered when considering the nature of mental properties.

The argument for physicalistic monism that can be derived from the premises of AM shows that the general claim that each mental event is identical with some physical event, but it does not say anything about particular or specific identities. How could particular identities be established? If PNCC is rejected, the question about particular identities is important also with respect to the question of whether mental events are physical events. One possible answer is that the only way to establish particular identities is to do science studying the brain. This strategy obviously depends on the assumption that mental states are brain states; they must be in the brain. Although this is the point that needs to be proven, we can accept the assumption for the sake of argument when considering whether identities could be established. That is, let us tentatively assume that mental states could be states of the brain. I believe an O-physicalist like Davidson or von Wright does not have convincing alternatives. The question now is whether this can be shown to be the case.

Token physicalism is consistent with genuine property dualism, and therefore token physicalism could be seen as a version of very weak form of physicalism. It nevertheless captures the idea that there is only one substance and that substance is physical. As Davidson notes, the “[…] important philosophical and, indeed, metaphysical thesis to which we are

\textsuperscript{648} Pereboom and Kornblith (1990) for example, reject the Humean account of causation and thereby also the token-identity of mental and physical events, offering instead the idea that mental states are physically constituted but not token-identical to the token physical states that constitute them.

\textsuperscript{649} Or whatever particular one wants to put here.
committed to is that mental events are identical with physical events”.  

It has not been shown that there could not be such “thing” as mental substance, but it has been shown, supposing that Davidson is right, that such substance could not interact with the physical substance and therefore it would be hard to become aware of its existence. It is clear that in the context of O-physicalism, ontological priority is given to the physical. One way to understand what this means is again in terms of supervenience. Davidson’s use of the terms “characteristic” and “determinination” in his discussions of supervenience has led some commentators to conclude that Davidsonian supervenience is:

[...] an inter-level metaphysical determination-relation that renders mental properties materialistically respectable…. The idea is that a reasonable materialism need only claim that physical facts and properties are the ontically basic ones, the ones that fix or determine all the facts.  

This description of Davidson’s position by Terence Horgan is similar to my interpretation, offered in sections 2.1.2 and 2.1.3, which suggests that AM is a form of robust ontological physicalism. It also gives support to the view of the previous section, according to which the way in which Davidson formulates the idea of supervenience is compatible with the idea that mental properties are physical properties; as Horgan notes, Davidsonian supervenience renders mental properties materialistically respectable. I agree with this and suggest, as was discussed in the previous section, that the view could be developed even further in the direction of materialism.

What is said about properties and supervenience applies also to supervenience and token-correlations. If the claim about supervenience is cashed out in substantial terms, it could be claimed that each mental entity must be made of physical entities. Notwithstanding the possibility of genuine emergence, this is one way to understand the claim that physical characteristics determine mental characteristics. This kind of compositional physicalism is one way to argue in favor of physicalism more generally. However Davidson’s relation to the claim that mental entities are composed of, but not identical to physical entities, is unclear. Although Davidson usually rejects the talk of “mental entities”, he nevertheless uses this expression when he claims that mental entities (objects and events) are identical to physical entities. What mental objects are remains a mystery, and I think it is questionable whether, given Davidson’s other views, the very term mental entity is a clarifying one. Objects and events are paradigmatic examples of entities, and the individuation criterion for both is their

---

650 Davidson, 1974b, 248, my emphases.  
651 Horgan, 1993, 565.
spatiotemporal location. Can mental entities be individuated in terms of this criterion? Let us for the moment ignore this question and conclude that the term *mental entity* is best understood as being a loose manner of speaking.\(^{652}\) It can be noted that whereas in “Mental Events” Davidson was discussing, as the title shows, the status of mental *events*, in his later writings he was not unwilling to talk about mental entities.

It seems to me that, given Davidson’s commitment to O-physicalism, we have reasons to think that mental entities are identical to physical entities in the sense of being composed of them. However, Davidson’s identity criterion for events is their identical spatiotemporal location, and it is not clear whether events generally can be analyzed into smaller components in any useful sense. In order to clarify Davidson’s identity-criterion, let us consider an event E. Since the problems that I want to raise are most obvious in the case of events that can be ascribed a mental description, we shall suppose that E is such an event.\(^{653}\) Let E be the event of Donald remembering where his keys are. A particular event E is now described in mental terms. Let us refer to event E as event M, since it now has a mental description and this is what “makes” it mental. Donald is aware of M in the sense that he has a linguistic thought “My keys are at work”. Perhaps it could be said that being in this state is the mental event M.\(^{654}\) According to the thesis of token-identity, M is identical with some physical event P. Let us follow Davidson and reformulate the problem in terms of descriptions. Instead of asking when an event M is identical with an event P, we shall ask when a sentence M = P is true. Davidson himself suggests that instead of asking the question of when are events identical, we should ask the general question of when a sentence of the form a = b, where “a” and “b” are supplanted by singular terms that refer to events, is true. He is thus looking for a general criterion of event-identity, and we already know what that is: events are identical if they occupy the same portion of space-time. Since we are interested in finding out whether *particular* token-identities can be established, we are asking the specific question when a sentence M = P is true.

Let us formulate the question in the following way. How to tell whether “Donald’s remembering where his keys are = activation of certain brain cells” is true? If Donald is aware of event E under description M, what are the reasons to think that description P (activation of certain brain cells) applies to E? What reasons do we, or Donald, have to believe that there exists a P (a spatiotemporal physical entity) to which description P refers? This Donald

---

652 The individuation criteria for mental entities will be discussed in sections 2.4.2 and 2.4.3.
653 Let us for the moment ignore the possibility that any event can be described as mental.
654 Perhaps the correct way to state this would be to say that being in the state of thinking “my keys are at work” is a mental event M and Donald being aware of this thought would be another mental event M₁.
certainly cannot know through introspection. The only event that Donald is aware of is M, which is a purely mental event, and of this event Donald can be aware through introspection. Putting neuroscientific speculations aside for the moment, what reasons do we have to believe that M is a spatiotemporal entity? It certainly does not seem to be such. In fairness to Davidson, I have to emphasize that the kind of approach suggested here would not make sense from his point of view, since the problem of token-identity is resolved simply by accepting PNCC. If this thesis is accepted, the problem of individuation becomes a question of how language identifies and sorts out events. But the individuation criteria for M are certainly different than the criteria for P; yet both descriptions are thought to refer to a single event E. The proposal that the problem of individuation is a question of how language is used cannot be an answer to the question of whether M is the same event as P; it cannot solve the question of whether descriptions M and P refer to the same event. Can we answer this question by saying that M and P are the same event if they occupy the same portion of space-time? It can be argued that if M and P are the same event, they occupy the same portion of space-time but, obviously, this criterion cannot be used if we want to know whether M and P are the same event. Suppose that M occurs “in” Donald and a neuroscientist observing a computer screen notices the occurrence of P. How to tell whether these seemingly distinct events are one and the same? With respect to this problematic question Davidson himself notes:

[…] perspicuous forms of the identity theory of mind require that we identify mental events with certain physiological events… for such theories to be interesting, there must be ways of telling when statements of event-identity are true.655

With the last sentence we can certainly agree. The criterion which Davidson first proposed for event identity was that events are identical if they have the same causes and effects. The inspiration for this criterion came precisely from a consideration of how brain events could be identified with mental occurrences. The suggestion was that what these events have in common is the sameness of their causes and effects. This criterion was later rejected by Davidson, although he seems to think that different ways to individuate events are possible and which one is chosen is, above else, a practical decision. But as the reference to identity theories shows, Davidson was concerned with the question of how to tell whether statements of event-identity are true. Let us now turn to the question of whether we can make sense of the claim that a specific mental event is a physical event, and whether it is possible to

655 Davidson, 1969, 165, my emphasis.
establish actual correlations or identities between these events. I will first briefly present von Wright’s views and shall then return to Davidson’s views.

2.4.2 A von Wrightian approach: Token-identity is a serious confusion

In section 2.2, I showed that von Wright expresses his sympathy for a physicalistic world view because of demands of “scientific intelligibility”. This being the case, we could assume that von Wright would accept also the truth of token-identity. But whereas Davidson is trying to provide an argument for such identity, von Wright’s general view is that the idea of token-identity is “a serious confusion”. In this sense Davidson’s contribution to physicalism is clearly more robust than von Wright’s, whose doubts about the coherence of the token-identity thesis can be seen as doubts about the coherence of the very position of physicalism.

On the one hand, von Wright claims that the suggestion that each psychological phenomenon has a neural counterpart contains “an important grain of possible truth…. The claim is delicate; token identity may be part of a possible truth. On the other hand, von Wright thinks that the very idea of token-identity is exceedingly obscure, unintelligible or confused. The reasoning behind these claims is that the empirical evidence suggests, and therefore scientific intelligibility demands, a defense of the view that there is something physical corresponding to every mental phenomenon; yet the result is a conceptual confusion when the idea of neural counterparts, the idea of “parallelism”, is coupled with the view that this parallelism amounts to identity.

Von Wright’s reasons for arguing that the idea of mental–physical identity is, if not complete nonsense, at least a serious confusion, are influenced by Wittgenstein’s views. The essential point is that there are certain requirements that the referents of the terms in true identity-statements have to meet, and these requirements are not met and cannot be met in the case of mental–physical identity. There do not seem to be non-problematic and acceptable criteria of identity between mental and physical phenomena. The phenomena in question, although they are the same in some sense, are individuated so differently that the identity-claim becomes absurd or false. It could be argued that since the identity of a phenomenon as such and such, for example as a mental phenomenon, depends on the way that it is individuated, the phenomena in question are so different that we cannot make sense of the claim that they are identical. Two major differences between mental and physical phenomena

---

656 Von Wright, 1989b, 22.
are the difference in their spatial location and the difference in the way that these phenomena are epistemically available.

These differences are closely related. A physical phenomenon, a brain state, exists in space. It can be observed, registered, and measured by an observer outside the brain. But a mental phenomenon, a feeling of pain or a belief that water is wet, is a “private property” of the person who has or experiences it. The fact that a brain state can be epistemically accessed from the outside whereas a mental phenomenon cannot be so accessed is one clear difference between them. This is so even though it is tempting to think that, in some sense, the phenomena in question are the same. Only the subject who is experiencing a mental phenomenon knows for sure whether the phenomenon exists at all: we feel that she is the final judge on the question of the existence of a mental phenomenon. This privacy of the mental must surely be taken seriously, because it is a clear feature which distinguishes these phenomena from physical ones. The problem of privacy does not occur in the case of a brain state. Anyone familiar with the relevant measuring technique can be certain that a specific brain state exists. The criteria for physical phenomena are public and such phenomena can thus be intersubjectively observed. From the fact that a mental phenomenon is private and thus not objectively accessible or observable, von Wright concludes that it is best to say that a mental phenomenon does not exist in space, and therefore it is absurd to identify the mental phenomenon with a physical phenomenon. We should not say that a mental phenomenon exists in space because there is no way in which we could show this to be true. The claim that mental phenomena could be found from the brain is nonsense, does not mean anything at all.

Von Wright admits that although this argument – which he considers to be the most important against the view that the relation between mental and physical phenomena is that of identity – must be taken seriously, it is also obscure and mystifying. Insisting on the identification of a mental phenomenon with a physical phenomenon would be to confuse things that have conceptually different natures. Whereas this sounds reasonable and the differences between mental and physical phenomena are real, the claim that mental phenomena are not in space is certainly mystifying. If mental phenomena are not in space, where are they? Should we conclude that the where-question itself is nonsense? I believe this conclusion would be too hasty, since we can understand the question of where mental phenomena are and, this being the case, the question is not complete nonsense. If mental phenomena were really not in space, if we could make good sense of this claim or perhaps even prove or show that mental phenomena have no spatial location, the claim of identity would be false and von Wright’s position would not be a version of physicalism. But, as I
have shown, von Wright subscribes to physicalism and rejects any form of substance dualism. Therefore the claim that mental phenomena do not occur in space should not be understood as a claim according to which these phenomena occur in a non-spatial, soul-like medium.

If we accept minimal O-physicalism, then we must accept that mental phenomena do not occur outside time and space. I think that what we should say, instead of claiming like von Wright that mental phenomena do not exist in space, is that the spatiality of mental phenomena is of a very different nature than the spatiality of physical phenomena. Both mental and physical phenomena are located in space, but a physical phenomenon has a specific location, which a mental phenomenon lacks. As I noted in section 2.3.1, an attempt to describe mental properties by modeling them in terms of physical properties leads easily to confusion. Likewise, to think of a mental phenomenon in terms of physical attributes, like spatiality, locality or extension, leads to a confusing picture because a mental phenomenon is not a thing-like entity. It cannot be intersubjectively observed and it does not have a clear spatio-temporal location, which means that the idea of its physical extension is nonsense. The lack of clear spatio-temporal location is true of many physical phenomena too, since a material object is a poor example of a physical phenomenon in general. But all physical phenomena are possible candidates for intersubjective observation. This is not the case with mental phenomena. The temptation to see mind as a thing-like entity, as a kind of object like the brain or some other organ, is what above all makes us ask questions about mind’s location.

Malcolm, a close friend of von Wright’s, argued that the idea of mind–brain identity, the idea that mental phenomena are token-identical with physical phenomena, is nonsense. It is somewhat surprising that von Wright does not mention Malcolm as an influence.\textsuperscript{657} Perhaps the most obvious reason for this is that Wittgenstein was an influence for both. Malcolm emphasizes that whereas a brain process occurs at a certain time in a certain location, we do not understand at all what the bodily location of a thought would mean.\textsuperscript{658} He claims: “I do not know what it means to say that a sudden thought is a brain process. In saying this I imply, of course, that the proponents of this view also do not know what it means.”\textsuperscript{659} Malcolm was challenging Smart’s position (who claimed that the identity thesis is an empirical hypothesis). Malcolm admitted that the accusation of Smart not knowing what he was talking about could

\textsuperscript{657} Von Wright does mention that Stoutland and Malcolm were his most active discussion partners after he got interested in the philosophy of mind in the 1980s. It is plausible to think that these philosophers influenced each other reciprocally.
\textsuperscript{658} Putnam has more recently defended this kind of view.
\textsuperscript{659} Malcolm, 1964, 171-172.
of course turn out to be false if Smart were, to Malcolm’s “surprise and gratification”, to explain his view. The fact that non-reductive physicalism is still a live position shows that such an explanation has not been given. This, that we cannot understand the meaning of identity claim, is what also von Wright says, but Malcolm made the point much earlier.\textsuperscript{660} Von Wright notes that he agrees with Malcolm about the unintelligibility of the view that mental phenomena are identical to physical phenomena, but concludes that they could never reach a complete agreement about these matters.\textsuperscript{661}

Malcolm claims that the bodily location of thought could have meaning only if we gave it a meaning by dramatically readjusting our language. We could adopt a new way of speaking, a new convention, according to which it would be perfectly natural to speak of the exact locations of thoughts. This could happen, but the new way of speaking would have to be justified somehow. Empirical evidence showing that thoughts could be arranged in specific location would be one good way to justify the new way of speaking. But as things currently are, the identity-claim is a claim without meaning. We often speak as if the location of thoughts would make sense, but on reflection it becomes obvious that this kind of talk leads to problems. Von Wright claims: “There simply is no ‘something’ which could conceivably be both a ‘brain state’ and a ‘mental state’…..”\textsuperscript{662} Could we in the future find something which could be understood as being both a brain state and a mental state? Von Wright seems to think that this cannot happen, because there is no “substrate” the traits (properties) of which the mental state and a brain state could be. Mental states are traits of a person, brain states are traits of the brain. But persons are not brains and brains are not persons. It seems to me that a person could qualify as the bearer of both mental and physical traits. But this says nothing about the identity of mental and physical phenomena; it merely shows that a person is a psycho-physical being.

How do these worries relate to Davidson’s position? He has suggested that events are identical if they have the same space-time location. It seems to me that this criterion cannot work for of the reason that von Wright and Malcolm state, namely that mental events have no extension in the sense that physical events have. How could we tell or show whether a mental event without a clear extension is or is not identical with a physical event, which presumably has a clear measurable extension? As I noted in section 2.4.1, what Davidson can say is that two events are identical \textit{if} they have the same space-time location, but this criterion does not

\textsuperscript{660} See for example, Malcolm, 1971 and 1986.
\textsuperscript{661} Von Wright, 1989b, 2002.
\textsuperscript{662} Von Wright, UPg, 14a.
answer the question of whether a specific mental event M (which a subject privately experiences) is identical with a specific physical event P (which an outsider observes). Davidson has no acceptable suggestion about how to answer the question of whether two events are identical or not. This has the unfortunate consequence that the identity-claim must be accepted, while accepting that we may never be in a position where the identities could actually be established. What we have to acknowledge, however, is that the problem of the identity-claim is not especially interesting from Davidson’s perspective, since from this perspective the truth of monism is already established. All events are physical, all physical events have a space-time location, and therefore also every mental event has a space-time location. But how does this reasoning fare against the von Wrightian claim according to which the spatiality of mental events is nonsense? It cannot answer this challenge.

2.4.3 Token reducibility

We have seen two different non-reductive approaches to the problem of token-identity. Davidson tries to defend the token-identity thesis and attempts to provide an acceptable criterion for event identity – whereas von Wright, while partly agreeing with Davidson, suggests that the idea of identity in this context is confusing. An interesting question arises: how should the possibility of actual token reducibility be seen from these perspectives? We know, for example, that a brain-tumor is a physical object which can be found from the brain. We know that a stroke is a physical event which can be observed occurring in the brain. Under the assumption that O-physicalism is true, mental phenomena are considered to be physical phenomena inside the brain. But can tokens of mental phenomena be actually found from the brain in a similar way as a brain tumor can? Is it possible to literally show that a certain brain state, a physical event, corresponds to a mental phenomenon? If a comparison between a belief and a brain tumor or between believing and stroke sounds odd, why is this so? They are all physical things – at least, so an O-physicalist would claim.

The question of the physical nature of mental phenomena can be approached from a philosophical (conceptual) perspective as Davidson and von Wright do, or from a scientific (empirical) perspective like many current philosophers and scientists studying the brain do. These perspectives are foreign to each other. Perhaps the main reason for this is the “crisis of philosophy” discussed in chapter one – the fact that the role of philosophy is becoming increasingly blurred as science keeps revealing new facts about reality. The obscure role of philosophy is especially evident in the philosophy of mind, where an important question – for
philosophy – is whether the problems are empirical or whether they can be solved through conceptual analysis. The question addressed in this section is that of what contributions a philosopher can make to the discussion concerning the possibility of actual token-reducibility, and more specifically, that of what Davidson’s and von Wright’s contributions to this discussion are. What is the attitude of a non-reductive physicalist towards token-reductionism and how does this attitude affect, if it does, the coherence of the position of non-reductive physicalism? The focus will be on Davidson’s views since von Wright’s skepticism about the idea of token-identity does not leave many reasons for him to think that mental phenomena could actually be reduced to physical phenomena. The views of Davidson and von Wright do not really clarify the empirical issues related to the problems of token-reduction, but their analysis of the relevant concepts is an important philosophical contribution.

The question about token reducibility is a question which is interesting from the perspectives of the various sciences studying the brain. In addition to its philosophical importance, it has consequences for the question of the extent to which the mind–body problem can be studied empirically and in an interdisciplinary fashion. The desire or need to establish mental–physical correlations is increasing as our understanding about the functioning of the brain increases. The problem of consciousness, a deep philosophical problem, suddenly seems to be a problem which can be solved empirically. Neuroscientists are making claims which would have sounded like science fiction just a few decades ago. Philosophy of mind is being integrated into cognitive science and the neurosciences. Non-empirical philosophical claims about the prospects or pitfalls of token-reducibility are facing empirical challenges to an increasing extent. There is clear empirical pressure on the kind of philosophy of mind which Davidson, von Wright, and other classical philosophers are doing. I think it must be admitted that the more we pay attention to the results of modern neuroscience, the more implausible certain parts of Davidson’s and von Wright’s views start looking. Perhaps certain aspects of their positions cannot simply be defended any longer; but even if this is the conclusion, I nevertheless try to clarify what kind of views they hold and why they hold them.

Many have interpreted Davidson’s view as suggesting that actual identities between particular mental and physical events could be established, that identities could be conceptually proved or empirically shown, and that it could be explained how or why a certain mental event is in fact a physical event (or how the causation between mental and the physical works). There is degree of vagueness involved in using expressions like “how”, and especially “why”, a mental event is a physical event. But we can nevertheless ask: why is the
neural basis of a given mental phenomenon the neural basis of that phenomenon rather than
another? It is known, for example, how the heart circulates blood and how (though less
clearly) a pain-killer reduces pain; the physical mechanisms are known. Can the alleged
physical mechanisms “behind” mental phenomena explained likewise? An affirmative answer
to this question would open the way for an empirical study of the mind through the study of
the brain. If Davidson is suggesting that an affirmative answer can be given, this would place
him firmly among those physicalistic philosophers who argue that the talk about mental
phenomena can be replaced by talk that refers solely to physical phenomena. This would
undermine Davidson’s position as a non-reductive physicalist, although it still would not
imply that Davidson would accept type–type identity between the mental and the physical.

The interpretation that Davidson is not denying the possibility of actual reductions is
justified, given that in “The Material Mind” Davidson claims: “[…] psychological events are
describable, taken one by one, in physical terms…”, and that “[…] for any particular, dated
psychological event we can give a description in purely physical terms….”663 Perhaps the
modal status of “can” in this expression should be analyzed more carefully, but one way to
understand it is to conclude that in Davidson’s opinion, physical descriptions of mental events
are possible in the sense that there is no reason to doubt that a science studying the brain
could not give a physical description of a mental event. But in order to give a physical
description of a mental event, one would have to know which physical event the mental event
in question is. How else could a specific mental event M be described as being a particular
physical event P?

Davidson does not propose to explain how a particular mental event could be
identified with a particular physical event, nor does he propose to explain how a description
of mental events in physical terms could actually be given. But he insists that there has to be a
physical description of any particular mental event. This must be the case, given physicalistic
monism. Davidson makes a strong statement when he notes: “Particular mental events can be
explained by physical science when we know particular identities.”664 This is a move towards
a view that mental facts can be explained in terms of physical ones, i.e. it is essentially a
reductive view. Whereas many philosophers, not to mention scientists studying the brain
would be happy to endorse this, it is not clear how well this view fits with Davidson’s overall
non-reductive position. But the claim that particular mental events can be explained by
physical science if their identities are known raises the question: how are we able to know, or

663 Davidson, 1973, 231, 249.
establish, these particular identities? How to tell whether a particular identity has been found? Without an answer to this question, the claim that mental events can be explained in a physical vocabulary, “by physical science”, is an empty prediction. Davidson himself seems to be ambivalent about the question of whether the establishment of identity is possible. On the one hand, he seems to think that empirical findings about the workings of the brain must be taken into consideration when philosophizing about the mind, but on the other hand he seems to put certain, quite strict, limitations on what science can reveal about the mind–body problem. I think this shows that in one sense Davidson’s position is anti-Wittgensteinian, or at least not as Wittgensteinian as his conception of philosophy might suggest. Wittgenstein was even hostile towards certain results of science, but Davidson seems to accept many results of science and even puts forward philosophical claims which could be interpreted as competing with empirical views. This reflects the general problematic position of contemporary philosophy of mind, since it, at least in the writings of Davidson and von Wright, tries to steer a middle course between being a wholly empirical and wholly conceptual discipline. This confusion in philosophy is partly a result of the naturalistic challenge it faces. In order to be a respectable philosopher of mind, one has to acknowledge what science tells us about the mind or brain. I believe that for philosophy this is not a healthy course, because it can lead to a situation in which philosophers start to think that their claims are open to the same kind of criticism that empirical claims are. In this work I shall not further consider this possibly dangerous threat to philosophy.

The view that philosophical claims should not be evaluated according to the same standards as scientific claims may as appear to be a form of philosophical arrogance. In “Mental Events” Davidson writes: “Do we, by declaring that there are no (strict) psycho-physical laws, poach on the empirical preserves of science – a form of hubris against of which philosophers are often warned?” He concludes that we do not, because whether we or not trust a philosophical statement is governed by empirical and theoretical concerns not indistinguishable from those of science, and goes on to say that in the question regarding the possibility of psycho-physical laws there seems to be no clear line between philosophy and science. This also is somewhat anti-Wittgensteinian and could be read as a counter-view to my claim, put forward in section 1.2.2, that philosophy and science should be kept clearly

---

apart and that this is Davidson’s view as well.\textsuperscript{666} I showed how Davidson claimed, for example, that “I am arguing, as philosophers will on \textit{a priori} grounds, that psychology and the social sciences are \textit{impossible}.\textsuperscript{667} In Davidson’s view philosophical claims are generally sensitive to empirical considerations; the reason that the claim of irreducibility of the mental is \textit{not} sensitive to empirical considerations relates to Davidson’s conviction that any reducibility of the mental would amount to a “change of subject”, that is, to a situation in which our concept of the mental would change. The various problems relating to this change will be thoroughly discussed in chapter three and in the appendix.

Davidson’s claim that mental events are recognizable and identifiable in the neurophysiological and physical realms when given neurophysiological or physical descriptions suggests that a mental event can be recognized and observed as a neurophysiological event. According to Davidson, discoveries clarifying the mental–physical relation can be anticipated and “understanding of correlations between events described psychologically and described in neurological terms will, and should, influence the philosophy of action.”\textsuperscript{668} Given the recent progress of neuroscience, this is perhaps a plausible claim although von Wright, for example, would be skeptical about the question of whether these discoveries are relevant for the philosophy of action and whether they should influence philosophy.\textsuperscript{669} But Davidson expresses also a more skeptical view about the prospects of science when he notes: “I have no hope that anyone will discover how to describe a state of mind or action in physical terms that are suitable for strict laws….”\textsuperscript{670} In my view this, a limitation of science, is a crucially important aspect of Davidson’s philosophy of mind. This aspect will be discussed in more detail in the appendix. There I will show that both Davidson and von Wright have a skeptical view about the possibility of science explaining in any significant way the mental life of humans. This skeptical view of Davidson’s also has consequences for the possibilities of token-reductions. Since physical terms which are suitable for strict laws are terms which are reducible to physics, the claim is that a state of mind cannot be described in terms which are reducible to physics. Since higher level physical sciences are, in principle, reducible to physics, the claim has to be that a state of mind cannot be described

\textsuperscript{666} On the other hand, with respect to the relation of philosophy and science, Davidson (1970, 215) notes: “Where there are no fixed limits only the timid never risk trespass.” This could be read as a claim that philosophers should not be too modest in their claims.

\textsuperscript{667} Davidson, 1964, 47, second emphasis mine.

\textsuperscript{668} Davidson, 2001e, 277.

\textsuperscript{669} At times Davidson seems to share this scepticism of von Wright, for example, when he is skeptical about the \textit{philosophical conclusions} that can be drawn from Chomsky’s work.

\textsuperscript{670} Davidson, 1999 p. 639.
in terms of any science which is reducible to physics. This, in turn, implies that mental events cannot be reduced to physical events.

According to Davidson, a systematic correlation between mental and physical events is forever out of the question, whereas particular correlations are, if not plausible, at least possible. Davidson argues that for a particular psychological event a description in purely physical terms can be given. But this view is repudiated to some extent when Davidson notes that: “Mental events are physical events, and so some physical description applies to each mental event. I don’t think I was saying anyone could say what the physical description is in particular cases, but if I did mean this I was wrong.” Davidson’s last word on the subject is surprisingly “pessimistic”. New scientific methods: “[…] may make it plausible that there is a physical event which is identical with each mental event.…” An interesting question is how scientific methods could make this sort of identity plausible. By being able to demonstrate the identity between a mental and a physical event? But how would such a demonstration fare against the von Wrightian challenge that identity is nonsense? In fact, despite the progress of scientific methods, Davidson nevertheless claims that “[…] they do not come close to really picking out a precise physical event, nor is it likely they ever will.” Indeed, picking out precise physical events that correspond to mental events is “a hopeless project” according to Davidson. But why is the project hopeless? It is curious that, according to Davidson, future scientific methods may make the token-identity thesis “plausible”. A philosopher with a naturalistic view would claim that these methods will make token-identity certain, and that picking out a precise mental–physical event will be anything but hopeless. On the other hand, it is also very curious that Davidson refers to new scientific methods as being arbiters in the question of mental–physical identity. The plausibility of a physical event being identical with a mental event should be a non-issue for Davidson, given his philosophical argument for monism. AM does not, or at least should, not need empirical support. Here Davidson’s oscillation between a position influenced by Wittgenstein and naturalism is again evident.

The problem with Davidson’s views about the possibility of token-reductions is that he often argues that there must to be token-identities while acknowledging that we may never find such identities. What is the rationale behind the claim that “it is possible to know that a mental event is identical with some physical event without knowing which one (in the sense

---

672 Davidson, 1999, 653.  
673 Davidson, 1999, 653.
of being able to give it a unique physical description)…”674 It is, of course, Davidson’s argument for monism, which we have found to be unsatisfying. As I have claimed in the preceding, it seems to me, however, that a non-reductive physicalist who accepts O-physicalism could accept also a view of mental properties or tokens according to which there have to be identities although we may never find them – and, as noted before, the acceptance of this claim would not make an essential change to his position. When Davidson asserts that particular mental events can be explained by physical science, this suggestion is lacking an argument which would lead one to believe that particular identities could be established. But the claim about the possible physical explanation of a mental event cannot stand without the view that identities can be established. The claim that mental events could be explained by physical science if and when identities are available is trivial, because then a mental event would literally be a physical event and there is no reason to think that physical events could not be explained by a physical science. I think that in order for the claim to be interesting, some kind of actual reduction from mental to physical is required. A philosophical argument that there are only homogenous events is certainly not enough, because it does not say anything about the mechanics of actual reductions. Since we are after particular identities, it is not enough to say, as Davidson does, that mental events in general are identical to physical events. Even if this is true, i.e. even if we accept Davidson’s argument for monism, the troubling question is: in the case of individual mental events, M, M₁,…Mₙ, with which physical events are they identical and, more importantly, how to show this? How to decide whether M is identical with P and M₁ identical with P₁ if P and P₁ occur at the same time but at different places? A “brain scan” example illustrates the point. How to decide which part of the scanned brain is identical with a mental event? If we take seriously von Wright’s claim that mental events cannot be assigned a location, the question of with which spatio-temporal region a non-spatial mental event is identical becomes extremely problematic. Von Wright’s claim certainly has some point and Davidson does not indicate how to solve this problem.

The problem of actual reduction is also an empirical problem. The human brain is a complex organ and countless processes take place in the living brain continuously. Paul Churchland, in a surprisingly pessimistic mood, claims for example that:

[...] your physical brain is far too complex and mercurial for its behavior to be predicted in any but the broadest outlines... faced with the extra-ordinary dynamical features of a functioning brain, no device

674 Davidson, 1970, 224.
constructible in this universe could ever predict your behavior, or your thoughts, with anything more than merely statistical success. The claim that no device constructible in this universe could ever predict my thoughts is remarkable. What justifies this skepticism? Again, before even trying to imagine how person’s thoughts could be predicted, we should ask a more moderate question: How to pin down correctly the physical processes which are M and M₁? Where should the boundaries of mental events be drawn, if the von Wrightian view that these events have no extension is accepted? In fact, if von Wright’s critique is accepted there is no way to draw the boundaries of mental events, because such events simply have no boundaries. If we agree with von Wright’s view that mental events have no boundaries, then it follows that they cannot be identical with defined regions of the brain. In this case, clinging on to the view that they are, or must be, identical with specific regions of the brain requires the acceptance of the very identity which, as for example Putnam has noted, is sui generis and without acceptable criteria. It is difficult to conceive of a mental event as being in some sense a non-spatial entity, but it is just as difficult to think of it as an entity with size and shape. Talk about “belief-boxes” is commonplace in the contemporary philosophy of mind, and pictures of brains in which beliefs and other mental phenomena are nicely located are familiar. Although all this talk and the graphic presentations of it can be thought as being partly metaphorical, they show that mental phenomena are often conceived as thing-like entities. To think of a mental phenomenon as a kind of object is a tempting picture, because if mental phenomena are real they have to be somewhere. On the other hand, it is difficult even to imagine a thing-like belief. If we agree with von Wright in his view that mental events have no boundaries, then it follows that they cannot be identical with defined regions of the brain. And then clinging on to the view that they are identical with specific areas of the brain requires the acceptance of the very identity which, as Putnam among others has noted, is sui generis and without acceptable criteria.

If mental events are actually non-reducible, this is obviously a more general problem for a physicalistic philosophy of mind and for the sciences which try to establish connections between mental and physical events. It is also a problem for research agendas which are based on the view that identity between the mental and the physical holds. John Dupre, a philosopher arguing against the idea of reductionism and emphasizing the disunity of science has noted: “[…] to assume that every event is amenable to physical explanation is wholly to

675 Churchland, 1995, 3.
beg the question in favor of physicalism." The problem is evident in the case of mental events. Why believe that mental events are amenable to physical explanations? The obvious but trivial answer is that mental events can be physically explained because they can be described in a physical language, and this in turn is a consequence of the fact that they are physical events. We have seen that the possibility of a description in a physical language has been used as a criterion that “makes” an event physical. The circularity is obvious: if events are physical because they can be given a physical description, and the reason why they are amenable to physical explanations is the fact that they can be given a physical description.

Even if we assume, for the sake of argument, that each event is a physical event, it may still be the case that they are not amenable to physical explanation – why insist that everything physical has a physical explanation? – but the philosophical worry here is more serious. Without actual token-reductions, what are the reasons for believing in the token-identity thesis? What is the reason for accepting that a mental event is a physical event? Doesn’t the impossibility of token-reductions open, in principle, the door for dualism? Davidson notes that if mental events and states are, or can be, physically describable one by one, then the issue is not ontological. In other words, dualism is not an ontological threat as long as mental events are physically describable. But what secures the physicalistic ontology in the absence of actual token-reductions? In answering this question, we could let science be our guide and follow the naturalistic or von Wrightian line, according to which demands of scientific intelligibility cannot really “tolerate” non-physical phenomena. With respect to the question of whether particular mental events could be reduced to physical events and explained in terms of them, we could wait for a scientific verdict. This was a common attitude among the identity theorists of the 1950s. If we follow those who claim that we already have various examples of successful mental to physical reductions and accept the view that this is a sufficient reason to believe that physicalism is true, then there seems to be no need for a philosophical argument for monism. But do we, or could we, have philosophical – that is conceptual – reasons to believe or disbelieve the thesis of token-reducibility and therefore the thesis of physicalism – which render them immune to empirical claims?

Davidson has claimed that in accepting physicalism we are committed to a philosophical thesis, and it is reasonable to expect that the expression “philosophical” must

---

676 Dupre, 1988, 49.
677 Davidson, 2001f.
678 Given that naturalism is sometimes termed as “scientistic” and von Wright is strongly opposed to a scientistic position, it is interesting that the argument for monism or the reason for defending it is the same in the case of von Wright and naturalistic philosophers: a demand of scientific intelligibility.
carry some weight in this claim. Everything comes down to this: if particular identities cannot
be established we can, following Davidson’s argumentation, believe that each mental event is
a general event E, and that these types of events are such that they can be given physical and
mental descriptions. But from this it does not follow that the relation between a particular
mental event M and a physical event P would be such that M could be given a physical
explanation, although M is P. The version of physicalism which follows if even token-
reductions are impossible is very weak and then, in the case of AM, the entire burden of proof
is based on PNCC, which I have shown to be highly suspect. In section 2.4.1, I proposed that
the particular contingent identities between mental and physical events can be established
only by doing science that studies the brain. This would imply that the question of the
possibility of token-reducibility is empirical. If PNCC is rejected and if it turns out that token-
identities cannot be established empirically, we have to ask what is left of token-physicalism
in the absence of empirical and philosophical arguments supporting it. From the perspective
of AM, the question of the possibility of token-reductions is uninteresting. The general thesis
of identity has already been established and no additional empirical evidence is needed for
monism. On the other hand, certain views of Davidson seem also to make this claim suspect.
We have seen how Davidson speculates about the possibility and nature of token-identities by
referring to the practices of current and future science. These speculations give the impression
that empirical evidence could be relevant for token-identity. But if this is so it undermines the
philosophical importance of AM and undermines Davidson’s original intent to provide a
philosophical argument for monism.

One interpretation of the supervenience-thesis that Davidson endorses seems to imply
that the relation between mental and physical events is similar to the kind of relation that
exists between a physical event and that event described in a higher level of a physical
vocabulary. On a physicalistic reading of AM a certain physical, say neural, event can be
described also in psychological language, which is subject to the principles of interpretation.
On this picture, there are no “ontological correlations” or “ontological relations” like
constitution or composition between events, since the mental description of an event is just
that – a description. In the discussion of token identity, all spatial terms are misleading since
there are no two events which could be related in a way which would imply that there is
spatial distinctiveness between the events. Constitution and composition, for example, can be
understood as referring to spatial relations; but there are no such relations in the case of
mental and physical events. In fact, what kind of relations could there be between events M
and P if they are the same event E? The mental description is supervenient mainly on
behavioral facts about the subject to whom the mental phenomenon is attributed, and therefore the question of whether this description reduces to a brain state of the interpreter is obscure.

If there are two persons A and B, it should be quite obvious that A’s description of B’s mental states cannot reduce to the brain state of B. If it reduces to anything, it is the brain state of A. However, if the problem is formulated in a somewhat more substantial sense and not from the perspective of person A, we can ask whether an existing mental state or event reduces to its physical “counterpart”. This raises the familiar worry of whether it makes sense to talk about counterparts if there is only one event being described. But even in order to formulate the worry in this way, it is already required that the mental status of an event is being thought of as merely a description. According to this understanding, an event E is a neural event (as well as physical, chemical etc.) and this event can also be described as mental, but there is no higher level entity which could be reduced to the neural event. There is no entity in addition to the neural event because the mental event is “nothing but” a neural event. This is a perfectly understandable physicalistic view, but does it make sense if it is considered in a broader setting? If it is claimed that a certain event E is a neural event and its also being a mental event is just a matter of us describing it as such, it seems to follow that the neural status of the event is somehow a more substantial fact about it than its being a mental event. But why believe this, and what could this claim even mean? If a person is aware of a certain mental event which takes place in them, it is obscure to say that this event is first and foremost or “nothing but” a neural event – thereby having an objective status as an event which can be observed from the outside. In a situation where person’s awareness of an event is an essential part of the event, there certainly seem to be two perfectly real aspects (mental and physical, or subjective and objective) of an event, and the question of what their relation to each other is can be asked. Can they be understood in terms of each other?

I think it is instructive to consider the issue in a more substantial way, one which clarifies the double aspect view. This may be in conflict with Davidson’s views about the nature of mental events but, as we will see below, the problems that are raised concern his position too. Moreover, the more substantial understanding of the problem fits well with common-sense intuitions and is therefore worth considering. In section 2.3, I claimed that the mind–body problem can be raised in a way that does not concern just the relation between mental and physical descriptions. To consider an example, let us suppose that I know that I believe at time t that “This table is pretty”. I have this thought “in my head” in the sense that I am aware that I describe to myself the table as being pretty. This mental process, my being
aware of what I am describing to myself, can be thought of as a true mental phenomenon. What is “mental” about this phenomenon is the subjective awareness involved, which cannot be wholly captured from the outside by an observer who is not experiencing this awareness. The outside observer cannot know for sure whether or not I am describing something to myself. Let us also suppose that at the time when I am having a thought, a neuroscientist is looking at a computer screen where the results of the scanning of my brain are being displayed. In this kind of situation we can ask: Is there something in my brain corresponding to my thought, or is some brain state my thought? This question is understandable, and in the context of O-physicalism in a physicalistic vein we are apt to think that the answer to this question is affirmative. We may even think that if I am really aware of my thought, something in my brain also has to correspond to my awareness. But can we say where my thought exactly is? Which slice of space-time is my thought?

What kind of a physical event a mental event is and where it is located are questions which do not arise, at least not in the same sense, when we are dealing with “purely” physical events or objects, i.e. with events or objects that cannot be given a non-arbitrary mental description. It is possible to say in great detail what kind of stuff water, for example, is and why it has the properties it has. If there is a glass of water and the water is heated, we can certainly say where the event of water’s warming up takes place, we can explain why the water boils when it is heated etc. Our knowledge about the connections between macroscopic and microscopic physical phenomena is impressive, and constantly accumulating. But we know next to nothing about the relation of mental and physical entities. If mental entities are physical entities then our understanding of these relations may increase in the future, but a serious problem remains: How can the subjective (the mental description) and objective (the physical description) aspect of an event E be brought together so that it would be possible to say that they refer to, or are part of, the same event E?

I have shown that Davidson has little to say about the prospects of actual token-reducibility and his final views about such a possibility are negative. His opinion seems to be that mental events cannot be clearly located, but that this is not a deep problem. According to Davidson, beliefs and intentions are not “little entities lodged in the brain”. This is a view shared by many non-reductivists like Rorty and Putnam. But what is this bold view based on? It is clear that the claim is not based on empirical evidence. But if non-reductive physicalists are not referring to empirical evidence, then what allows, for example, Rorty to say: “[…] we do not expect neurology to provide a fairly detailed microstructural story about any given mental state. We do not expect to be able to change people’s beliefs and desires by tinkering
with their brains…”?679 By “we”, Rorty must mean mainly philosophers – and especially non-reductive physicalists – because it is implausible to think that scientists studying the brain would have this attitude. Many of them would be willing to say that what we expect is \textit{precisely} that neurology will provide a detailed microstructural story about any given mental state.680 In fact, research on the neural correlates of individual thoughts is being done constantly.681 It is interesting that on the one hand such research is being done and at the same time various philosophers like Malcolm, Putnam, Rorty, Davidson and von Wright claim that such research is basically impossible. This situation says something revealing about the current relation between philosophy of mind and the sciences of mind. The previously mentioned debate between Hacker / Bennett and Dennett / Searle is a concrete example of the complex nature of the disagreement and also of the passion which surrounds the discussion.

Given that current empirical evidence is apt to suggest otherwise, what is Davidson’s and Rorty’s reason for claiming that beliefs are not little entities in the brain? Ignoring here the fact that Davidson does not clearly define the expressions “state” and “entity”, perhaps the best way to understand the view of Davidson and Rorty is to realize that their understanding of the nature of beliefs, as \textit{mental} entities, prevents them from seeing beliefs as independent physical atoms of the brain’s architecture. The holistic nature of the mental is the main reason for this. As Rorty argues: “If we assert token–token identity between sentential attitudes and neurological states, the latter will have to be states of an entire nervous system, rather than nicely specifiable bits of that system which could be zapped in order to alter beliefs and desires one at a time.”682 This claim is the consequence of the view that, because of the holistic nature of thoughts, a specific thought cannot be changed without changing the relations between it and many other thoughts. Many philosophers, most famously Davidson and Putnam, have argued that the holism of the mental has the consequence that token-identities obtain between sentential attitudes and the states of an entire nervous system. The identities cannot obtain at a more local level. In my opinion it is unclear how seriously Rorty’s expression “entire nervous system” should be taken. Literally speaking, the human nervous system does not consist of the brain alone, since the spinal-cord is also part of the central nervous system. Is Rorty suggesting that sentential attitudes are token-identical with neurological states which are the total states of the brain and spinal-cord taken together? This would mean that my belief would be partly located in my spinal-cord! This conclusion shows

680 For this kind of view see for example, Bickle, 2003.
681 For a general discussion see Metzinger, 2000.
the problems that become evident if an expression like “the entire nervous system” is used loosely and if non-empirical claims about the assumed identity are being made. It is of course possible to settle for the view that my belief is token-identical with a state, or a space-time region, which includes a part of my spinal-cord. This would mean that my beliefs would literally be outside my head. This conclusion would agree with von Wright’s and Malcolm’s views. They would claim that a view according to which my belief is identical with a part of my spinal-cord is as nonsensical as the view according to which my belief is identical with a part of my brain.

Davidson and Rorty are not the only philosophers who see holism as an obstacle for specific token-reductions. Putnam has also suggested that holism prevents the possibility of specific mental–physical identities and even token-correlations. If there cannot be token-correlations, a token-reduction is also obviously out of the question. What it would mean to claim that thoughts were independent and nicely specifiable bits of the brain? According to those who emphasize the holism of the mental, it would mean that we could take one belief, say “Dogs are animals”, in isolation and perhaps even change it, without changing (any of the other) physical relations in the brain. A more radical conclusion considered by Putnam is that if propositional attitudes were clearly specifiable bits of the brain it would be possible, in principle at least, to isolate a specific thought and have that single thought stored in a test tube in a laboratory. In Putnam’s view this idea is utterly crazy. Perhaps it is too easy to agree with Putnam. An isolated thought sounds nonsensical because its content could not be identified or defined without relating it to other thoughts. What would an isolated thought be about?

The suggestion of Davidson, Putnam and Rorty that holism creates serious problems for the possibility of token-reductions is interesting because ontological conclusions are being drawn from a consideration about the peculiar features of mental states and their attribution conditions. We can agree that holism is a problem, but should we perhaps go even further than Putnam in his claim and say that, because of holism and externalism, it does not make sense to think of beliefs as internal states of the brain, or more broadly, of the body at all. This is the von Wrightian–Malcolmian view – the conclusion being that it is not clear what the mind-brain identity amounts to or could amount to. This is not as radical a conclusion as it may sound; the mind–body problem is still unsolved and it is not obvious that the physicalistic answer or the “mind–brain identity thesis” will triumph. Those who keep

---

identifying the mind with the brain ignore the non-reductivist challenge, according to which it simply makes no sense to say that mental phenomena are identical with the states of the brain. If this is the case, then claims about the brain should be withheld when talking about mental phenomena. This is a conclusion drawn by von Wright and Malcolm, and it is understandably not a well accepted view in the contemporary philosophy of mind.

Are Davidson and other non-reductivists giving a satisfying argument that shows actual token-reductions are not possible? The thesis of the holism of the mental gives rise to one such argument. Although holism is an obstacle for successful token-reductions, there is nevertheless the following problem: It is true that the content of thoughts depends on the contents of other thoughts. But it seems to me that when we are talking about thoughts we are already, so to speak, in the realm of the mental. Does it make sense to say that the physical relations between the states of the brain could constitute the identities of brain states in a similar way as in the case of mental states? Presumably not, for holism is a feature of the mental realm. But can this form of holism then preclude actual token-reductions between mental and physical states? As will become evident in chapter three, an important reason for defending the irreducibility of the mental is precisely the conviction that certain relations which obtain at the mental level do not obtain at, and cannot be transferred to, the physical level. Connections which depend on the propositional contents of thoughts cannot exist in a vocabulary which is devoid of these kinds of contents; these connections do not exist in a vocabulary where the contents have no relevance. It does not make sense to speak about logical relations between physical events. But if this is the case, can Davidson and Rorty really argue that the holism of the mental is the reason that there cannot be locally identifiable physical states which are the physical “counterparts” of specific mental states? Can the holism of the mental be used as an argument against reducibility if holism has no relevance at the physical level? Why think that the holism of mental states would lead those states to be global, not local, states of the brain? Although it may be the case that in order for a person to have the concept of a “cat”, the person must have many other concepts, it is not at all clear that these conceptual connections are, or must be, realized in the brain in a way which prevents, or would prevent, actual token-reductions. Given that it is not well known how the brain functions, it seems premature to say that token-identities between mental states and the brain exist only at the level of mental states and the whole nervous system.

The main question is whether the holism of the mental has ontological implications if our ultimate ontology is physicalistic. Does the holism of the mental have any consequences for the question of what kind of physical phenomena mental phenomena are? It is interesting
that Davidson thinks that he agrees with Rorty in the view that the holism of the mental does not have ontological implications. But it seems to me that Rorty draws precisely ontological conclusions from the fact that mental states are holistic by nature: he believes that these states must be states of the whole nervous system. This is a specific claim about the physical nature of beliefs. What Davidson must mean is that he agrees with Rorty’s view that holism is not a threat to the reality of mental states. Holism does not, pace Fodor for example, threaten the ontological status or existence of mental states. But it is unclear what Davidson’s main reason for rejecting the possibility of establishing psycho-physical token-identities is. Since he argues in favor of physicalism on a priori grounds, he has to think that there must to be such identities. Insofar as beliefs are real physical phenomena, there must be spatiotemporal regions which could, in principle, be identified as beliefs. As we have already concluded, conscious belief – as long as it is a physical phenomenon – must be somewhere, and there is no argument whatsoever in Davidson’s writings which would offer support for the claim that this belief would be identical with a state of the whole nervous system. For Davidson, the claim that mental phenomena are token identical with physical phenomena is not an empirical claim although the question of “how beliefs are physically constituted” is to be decided empirically – and for this question Davidson offers only negative, skeptical and obscure answers.

Given that the physical constitution of beliefs, insofar that this expression makes sense at all, is an unresolved empirical question, why should we think that mental states correspond to global states of the brain? What kind of states are these “global” states in the first place, should the spinal-cord or indeed the whole peripheral nervous system be included in the “physical base” of beliefs? It is a contingent empirical fact that, most likely, I would not have thoughts if I lacked a central nervous system. But, to elaborate further, I would not have thoughts if I lacked a heart. Should we conclude from this that if we assert token-identity between sentential attitudes and physical states, the latter must include my heart as a part? The exact location of thoughts may be an empirical question, but the philosophical suggestion that they are identical with the states of the whole nervous system is not obviously better than the suggestion that thoughts are identical with the states of the heart.

I believe that Rorty and Davidson are not aware of the exact details of how the nervous system, or the brain, functions. If they are aware of these details, they do not use this knowledge to support their view concerning the location of thoughts. Many philosophers, von Wright among them, openly admit that they do not know at all how the brain functions. Yet these philosophers are willing to make claims about the mind–body relationship that are not
entirely conceptual. This is one reason why the dialogue between scientifically-minded philosophers and those who are interested in conceptual arguments is often so futile. We can assume that when Davidson or Rorty are speaking of “global” and “local” brain states, they do not have anything very specific in mind. Is Rorty’s claim – that propositional attitudes are states of the “whole nervous system” is unlikely – based on neurological evidence? Likely not. But is it then meant to be a claim about the physical structure of the brain which could actually and seriously compete with empirically based neurological claims? Scientists studying the brain would be skeptical of these kinds of a priori philosophical views. They would perhaps judge them as irrelevant and consequently would not take them seriously. Whether this in turn moves philosophers much is questionable. Rorty, Davidson, Putnam and von Wright are not among those philosophers who try to reconcile philosophy and science. Their attitude is totally different from those who are openly doing neurophilosophy. The view of, for example, Paul Churchland is based, partly, on what we already know or seem to know about the brain whereas the view of Davidson, von Wright and Rorty are based on what we know or seem to know about the mind. Since we seem to know much more about the mind than about the brain, we may be easily tempted to draw conclusions about the latter from the former. But as those studying the nature of the brain know, this a priori strategy does not often work very well. I think that when Davidson and von Wright offer speculations about token-reductions they are not being true to their Wittgensteinian conception of philosophy. In this respect they are being pseudo-naturalists whose claims cannot really compete with empirical claims. Whereas I think we should express great doubts against scientistic naturalism, we should also be aware of the dangers and temptations of a priori scientific philosophy.

What, in sum, can be said about the non-reductivist’s arguments against actual token-reducibility is this. It is obvious that the conclusions of Davidson, Putnam, von Wright or Rorty are not a result of their empirical study of the brain. These philosophers seem to hold the view that empirical study of the brain cannot find beliefs in the brain because, needless to say, beliefs are something else than specific little entities in the brain. They are either states of the whole brain or states of the whole person. But it is quite unclear what the talk about entities, states, processes etc. mean in this context. Are philosophers and scientists talking about the same things when they use these terms? Most scientists studying the brain think that whether or not beliefs can be found from the brain is entirely an empirical question. Perhaps

---

685 For a general view see Churchland, 1995.
beliefs are not “entities” or “objects” in some philosopher’s obscure sense of the word, but a scientist who is actually studying the brain could be convinced that beliefs correspond to local states or processes of the brain – and empirical evidence may support this view. A scientist would not problematize these issues in the same way as a philosopher does. The consequence of this is that, from the scientific perspective, the a priori arguments of philosophers insofar as they attempt to compete with empirical views look suspicious and are accorded little relevance.

It is plausible to think that philosophers of mind who are not “scientifically oriented” understand the nature of beliefs very differently than neuroscientists who study the brain do. This is a reason why philosophers so often complain that scientists “have missed the point”. On the other hand, in his recent book the philosopher / neuroscientist John Bickle claims that many philosophers of mind are just not aware of the current situation in the sciences, and therefore many of their claims can be refuted by empirical evidence. Bickle has argued that scientists are actually doing the kinds of reductions the possibility of which is being denied by certain philosophers. On the other hand, those philosophers who, for example, want to defend the autonomy of psychology insist that Bickle’s claims are simply untrue. This is yet again an interesting fact about the sociological trends in contemporary philosophy. Bickle’s view is nevertheless shared among many empirically oriented philosophers. Patricia Churchland, for example, claims that many philosophers are still under the impression that philosophical questions could not be solved empirically. Against these views we have the recent claim of Bennett and Hacker, who insist that the task of philosophy is not to put forward claims subject to empirical testing. The task of philosophy is a clarification of existing views in terms of conceptual analysis. I believe this is an important remark on the nature of philosophy, and as the debates show the question about the nature of philosophy does have concrete consequences for the kind of philosophy that is being done. It seems that many of Davidson’s suggestions about the prospects of token-reducibility amount to no more than speculation, which is especially dangerous if it is meant to compete with empirical views. Philosophers should not provide these kinds of suggestions, according to one plausible view, concerning the nature of philosophy. Churchland’s claim that philosophers are “under the wrong impression” is a non-starter for a philosopher like von Wright, who has a specific and different view about the nature of philosophical problems.

687 Churchland, 1986.
Here we again encounter the crucial problem which I introduced in chapter one, namely that philosophy in general and philosophy of mind in particular are facing a crisis, since their conclusions are becoming irrelevant in the light of empirical evidence. The discussion about the possibility of actual token-reductions from the viewpoint of non-reductive physicalism shows how obscure some of the arguments, such as Rorty’s claims about the relationship between “sentential attitudes” and “the nervous system”, are. The serious threat for any non-reductive philosophy of mind is simply that the philosophical claims of this position will not be taken seriously in the future. Bickle, for example, is already explicit in his view that scientists should not especially care about the arguments of philosophers, and Paul Churchland’s view that philosophical problems are empirical problems leaves no room for speculative philosophical arguments.

Whether or not philosophical arguments have relevance for science is an important question, and philosophers should be cautious in their claims especially in the areas which are becoming increasingly empirical. Philosophy of mind is such an area. I think that the position of non-reductive physicalism, insofar as it tries to hold on to the non-reducibility of mental phenomena, should present an argument which would show the impossibility of actual token-reductions, and this argument should be such that it is immune to empirical counter-examples. Since the prospects of coming up with this kind of argument are not particularly bright, non-reductive physicalism should perhaps settle for the next best thing: to critically evaluate scientific claims without putting forward a theory of its own. I think that von Wright has succeeded better in this second task than Davidson, who constructed a theory of the mind. AM has its merits, but the details concerning the ontological mind–body problem which this position suggests certainly look suspicious from the naturalistic perspective.

2.5 The nature of mental phenomena

I have described the nature of the physicalism that Davidson and von Wright defend, and the reasons why they defend it. We have seen the arguments these philosophers give for their physicalistic positions. We have also seen what their view about mental properties, token-identities, and token-reductions is. These are the main questions that are usually raised in the discussion about the coherence of non-reductive physicalism. With respect to all these issues, Davidson’s and von Wright’s views are somewhat obscure but quite similar to each other. Despite the obscurities, I believe we have arrived at a view of what Davidson’s and von Wright’s understanding of the physical nature of mental phenomena is. Whether it fits with
empirical facts is doubtful. Some of the claims of von Wright and Davidson certainly go beyond “describing features of concepts”, and as such face the challenge of being refuted empirically.

Because the position under discussion is non-reductive physicalism, it is to be expected that its conclusions about the physical nature of mental phenomena are inconclusive. It would be surprising if a non-reductivist had many specific views about the physical nature of mental phenomena. A non-reductivist cannot offer a detailed view of the physical nature of mental phenomena if his position is to remain coherent. Specific claims about the physical nature of mental phenomena would be almost like contradicting the basic position taken, according to which not much can be said about the relationship between the mental and physical. As Davidson notes: “I find no plausibility in the idea that thoughts can be nomologically identified with, or correlated with, phenomena characterized in physical or neurological terms.” 688 In order to hold this view too much need not be said; but how should a Davidsonian position react to current empirical results which suggest that such nomological correlations are possible? Is Davidson’s understanding of nomologicality perhaps too strict or restrictive? 689 A non-reductive physicalist cannot provide any details about the physicality of mental phenomena if he is to insist, as Davidson does, that: “No physical or non-mental science could be expected to explain thinking, the formation of intentions, or the states that characterize our mental lives and explain our actions.” 690 Given these “convictions”, the physical nature of mental phenomena has to remain incomplete.

But what can be said about the mental nature of mental phenomena? A substance dualist would perhaps say, for example, that mental phenomena do not occur in space. Mental phenomena would thus have at least one important property that physical phenomena do not have. Non-reductive physicalism is accused of being a form of “property dualism” (according to which the brain has properties possessed by no other kind of physical object), but as I have argued, this is not an illuminating way to describe the position. Even a reductive physicalist would admit that the brain does have properties possessed by no other kind of physical object: it is for example conscious. As far as we know, other material objects are not conscious. The view that brain is a special kind of material object does not imply a substantial form of dualism. Although a “property dualist” need not commit himself to a robust form of dualism,

---

688 Davidson, 1982, 100.
689 This question will be addressed in detail in chapter three.
690 Davidson, 1990b, 92.
it can nevertheless be asked what the properties of mental states are that distinguish them from physical states. We can ask what kind of phenomena mental phenomena are.

Davidson and von Wright claim that mental states are physical states under a certain description. There is therefore no “ontological mystery” about the nature of these states. This is explicitly Davidson’s view, whereas von Wright’s view is more wary. Given monism, there cannot be any issue about the ontological status of mental entities, as Davidson himself claims. In my view this claim is highly questionable, but it shows Davidson’s view of the mental. The exact physical nature of mental states is currently unclear; these states are neural states, and the question about their physical nature is left for neuroscience to answer. To describe phenomena as mental is to say that they can be described in a certain vocabulary. The mind–body or mental–physical distinction is not an ontological, but only a conceptual distinction. But what are the essential properties, if such there are, of mental phenomena? A non-reductive position emphasizes, so to speak, the mental aspect of mental states. What do Davidson and von Wright say about this aspect? The details they give are actually few. What, then, is the merit of their positions as non-reductive positions? It could be argued that they say what mental phenomena are not, instead of saying what they are. An essential question to ask of Davidson’s and von Wright’s positions should be: how is the nature of mental phenomena defined or clarified? Perhaps it could be claimed that their respective philosophies of mind are attempts to come up with such a clarification. But their detailed views about the mental nature of mental phenomena are scarce. This is unfortunate, because although there may not be a question about the ontological status of mental entities in the context of monism, I think there is nevertheless a kind of ontological question to be answered, namely the question about the “mental mode of being” of mental entities. Given monism, there is no question about the ontological status of gases or liquids, but there is, for example, a question about how they differ from solid objects. Likewise, one should arguably be interested in the peculiar general features of mental phenomena, because understanding these features helps to better understand the human position within a physical reality.

In my opinion, it is not enough to say that being amenable to a mental description in a certain vocabulary makes a phenomenon mental, because the consequence of this would be that any event would count as mental if given a suitable description. This does not go well with our actual, common-sensical understanding concerning the status of mental phenomena. In the contemporary discussion, references to mental states are made without clarifying what

---

691 Davidson (1970) admits that his position has this consequence.
the essential features of these states are. A similar problem was encountered in our discussion of mental properties: it is doubtful whether there are the kind of mental properties which most contemporary philosophers of mind take there to be. With respect to tokens of mental phenomena, the essential problem is this. Let us suppose that person A attributes a mental state to Georg. He says: “Georg believes that he is late for the meeting”, thereby attributing to Georg the belief that he is late for the meeting. This belief is a mental state. Seen from the perspective of physicalism, the belief could be seen, for example, as a physical disposition. This is how Davidson sees the nature of belief: “Anomalous monism makes sense of the claim that attitudes are dispositions to behave in certain ways, which in turn are physiological states, which finally are physical states....”\(^{692}\) This, incidentally, is in rough outline also Quine’s view about the nature of belief. Davidson seems to accept Quine’s view that beliefs are dispositions which finally are microphysical states of physics, although it was Quine who, after rejecting his eliminativist leanings, came to endorse AM as the most plausible theory of the mind. It is a plausible thought that Davidson adopted the general view concerning the physical nature of mental phenomena from his teacher, who then followed Davidson’s further reasoning and finally accepted the inevitable irreducibility of the mental. The kind of transition which can be seen in Quine’s views is nevertheless absent in Davidson’s writings. At times Quine described his position as reductivist, while at times his position was best interpreted as a form of eliminativism. Davidson, on the other hand, was always a non-reductivist and the only transition that can be detected is a move towards a position which more strongly emphasizes the irreducibility of the mental.

A brief description of Quine’s position is useful because Davidson has often been interpreted to hold the same position as Quine, i.e. continuing a strict eliminativist approach in the philosophy of mind. The critical Quine of \textit{Word and Object} famously noted:

\begin{quote}
One may accept the Brentano thesis either as showing the indispensability of intentional idioms and the importance of an autonomous science of intention, or as showing the baselessness of intentional idioms and the emptiness of a science of intention. My attitude, unlike Brentano’s, is the second.\(^{693}\)
\end{quote}

The main reason why Quine distinguishes himself from Brentano’s view is that “[...] the essentially dramatic idiom of propositional attitudes”\(^{694}\) 
\textit{has no place in serious science} and that the “[...] mentalistic vocabulary is stubbornly at variance with scientific patterns”.\(^{695}\) All

\(^{692}\) Davidson, 1997c, 72. See also Davidson, 1990b. Von Wright (UPg, 1) agrees: “To believe is dispositional.”
\(^{693}\) Quine, 1960, 221.
\(^{694}\) Quine, 1960, 219.
\(^{695}\) Quine, 1975, 92.
things considered, “The propositional attitudes are in a bad way”\footnote{Ibid.} and “The easy familiarity of mentalistic talk is not to be trusted”.\footnote{Quine, 1975, 95.} This is Quine’s most critical phase. He notes: “The bodily states exist anyway; why add the others?”\footnote{Quine, 1960, 264.} States of mind can be ascribed to the body, which leads to the view that “The mind goes by the board, and will not be missed.”\footnote{Quine, 1985, 5.} Quine seems to be ambivalent about the question of whether mental phenomena should be reduced to physical phenomena or eliminated in their favor.\footnote{Quine has often been interpreted as one of the main influences behind eliminative materialism. But these interpretations should take into account Quine’s view that: “Some may... find comfort in reflecting that the distinction between an eliminative and an explicative physicalism is unreal”.\footnote{Quine, 1975, 95.} 700} A physicalistic conclusion is: “I end up with the so-called identity theory of mind: mental states are the states of the body.”\footnote{Quine, 1979, 163. See also Quine, 1987, 133 and 1995a, 87.} Surprisingly, Quine is not expecting a reduction of mental to the physical. He says:

> It [my approach] to physicalism is not a reductionist doctrine of the sort sometimes imagined. It is not a utopian dream of our being able to specify all mental events in physiological or microbiological terms. It is not a claim that such correlations even exist, in general, to be discovered; the grouping of events in mentalistic terms need not stand in any systematic relation to biological groupings.\footnote{Quine, 1975.} 702

On the other hand, in “Mind and Verbal Dispositions” Quine distinguishes between three levels of explanation, the mental, the behavioral, and the physiological.\footnote{“Our three levels [mind, behaviour, neurophysiological] thus are levels of reduction: mind consists of in dispositions to behavior, and these are physiological states.” (Quine, 1975, 95) Moreover, according to Quine, in explaining mentality those behavioral dispositions whose physiological mechanisms seem “most likely” to be detected in the foreseeable future should be favored. I take it to be rather mysterious how this view should be interpreted. A hint is given when Quine (1960, 225) writes: “The farther a disposition is from those that can confidently be pinned on molecular structure or something comparably firm, the more our talk of it tends to depend on a vague factor of ‘caeteris paribus’”. This suggestion seems to indicate that molecular structure, or something as “firm” or “real”, would be the mechanism to look for.} According to Quine, the mental hardly deserves to be called an explanation, whereas the physiological is the deepest and most ambitious way of explaining. The three levels are levels of reduction.\footnote{Quine, 1975.} 704 Like in Davidson’s view also, mind consists in dispositions to behavior, which in turn are reduced to physiological states. I think that Davidson’s suggestion that AM makes sense of these levels of reduction could be interpreted as a suggestion that AM clarifies Quine’s original position.

The conclusion at which Quine arrives is the identity theory of mind: mental states are the states of the body. But, right after praising the identity theory, Quine becomes cautious and warns that the theory has its problems. What Quine ends up holding is the claim that “[...]
instead of saying that mental states are identical with physiological ones, we could repudiate them; we could claim that they can be dispensed with, in all our theorizing, in favor of physiological states.... He thinks that the identity theory may excuse a “recourse to mentalistic semantics”, but that the repudiation theory blocks this option and should therefore be favored. An identity theory could be “abused” by a philosopher who would like to excuse his free and uncritical use of mentalistic concepts. The virtue of the repudiation theory is that it precludes this kind of abuse. However, later Quine goes through considerable change of mind. He claims:

There is no mental substance, but there are irreducibly mental ways of grouping physical states and wants. The keynote of the mental is not the mind; it is the content-clause syntax, the idiom ‘that p’. Brentano was right about the irreducibility of intensional discourse.... Its irreducibility is all the more reason for treasuring it: we have no substitute.

This is a remarkable change of mind, and the exact reason for it is unknown. It may relate to Quine’s doubts about the capacity of physics to describe “what there really is”. In the 1960s Quine thought that Brentano was wrong; later he accepted that Brentano’s view was, in some central respects, correct. Brentano was correct in suggesting that our intensional vocabulary is irreducible and indispensable. Whereas in 1960 Quine couldn’t find a “base” for mental talk, in 1995 he concludes that “[…] the conspicuously intensional idiom of propositional attitude finds it niche.” Perhaps surprisingly, the final position of Quine is the position of AM. Here we have an interesting example of how the “master learned something from the acolyte”. Quine’s last word is: “The age-old duality of mind and body has not dissolved; it has shifted from substance to concepts... as such it remains irreducible and “[...] we must... acquiesce in the psychophysical dualism of predicates” and “Eventual reduction of [propositional attitudes] is hopeless... in respect of cognitive content they are danglers...” This being said, I think it must be emphasized that, despite Quine’s acceptance of conceptual

---

705 Quine, 1975, 94, my emphasis.
706 Quine, 1990, 71, my emphasis.
707 Quine, 1995, 98.
708 See, for example, Quine, 1985, 7, 1987, 133, 1995, 87 and 1995b, 358. We can note, as the reference to 1985 article shows, that Quine’s move towards anomalous monism was not especially recent. Woods (1992, 559) perspicuously notes: “Quine succumbs to a kind of dualism, an uncongenial fate for one who aspires to monism.” Of course, the same could be said also of Davidson and von Wright.
709 Maybe the seeds of non-reductionism were already planted in 1960, when Quine comments that his approach is less reductive than Frege’s version of number and acknowledges Davidson as one of those who made him to see this. (Quine, 1960, 265 fn. 7)
710 Quine, 1987, 133-134.
711 Quine, 1995, 87.
712 Quine, 1995b, 358. It is an interesting question what Quine means by “danglers”. This term could be expected from a philosopher who defends emergent properties, but it is hard to think that Quine would be one of them.
dualism he is best interpreted as a *pseudo-anomalous monist*. Until the end, even after becoming an anomalous monist, he hoped that psychology could be reduced to a certain kind of materialistic view and encouraged the efforts of, for example, the Churchlands to “reclaim territory from the intensional side”. A “true” anomalous monist would not conclude, as Quine does, that “Mentalistic words contribute vitally to everyday explanation and are practically indispensable, today anyway….”

After this consideration of Quine’s view about the physical status of mental phenomena, we should ask what kind of state a mental state is, *mentally* speaking. What is mental about a belief? Do we, by saying that Georg believes that he is late for the meeting, imply that Georg is *aware* of his belief? If not, it seems that nothing mental or physical needs to be “in” Georg at the time of the attribution. We say that people can “have” all kinds of beliefs although they never *believe* those beliefs. On this account, mental states like beliefs are physical dispositions of which a person needs not to be aware. Yet we can correctly attribute these states to the person. In fact, we may be in a better position to attribute these beliefs to a subject than the subject himself. I believe this raises very interesting questions about the possibility and nature of unconscious mental states. However, if nothing mental or physical is “in” Georg at the time when a certain belief is attributed to him, this means that the ontological status of mental states is suspicious and their mode of being is strange. The existence of mental states would be tied to the attributions of such states. A physical disposition would count as a mental state only because it would be described as being mental. This would cause the mental to be in the eye of the interpreter. According to some, this would mean that the very existence of mental states would be the result of a process of interpretation. This seems implausible. It would come close to a form of antirealism about beliefs, i.e. a form of eliminativism. The flaws of such a position I have already discussed.

---

713 That Quine welcomed a reclamation project and at the same time saw his own approach as a more cautious one can be seen already in Quine’s earlier writings. In *Word and Object* Quine notes: “The radical reduction that would resolve the mental states into the independently recognized elements of physiological theory is a separate and far more ambitious program [than Quine’s reductive approach].” (Quine, 1960, 266) On the other hand, when it comes to such radical reduction, “I should prize that achievement, and I should regard nothing short of it as a full causal explanation of the mental states and events in question….” (Quine, 1977, 637) This is interesting, since it seems to imply that the “full causal explanation” takes place at the most basic physical level. This view is relevant for the discussion of mental causation, which will be addressed in chapter four.

714 Quine, 1999, 409, my emphasis.

715 Fodor, among others, raises this criticism against Davidson.
2.5.1 A distinction between conscious and non-conscious beliefs

If the conclusion that mental phenomena are in the eye of the interpreter is to be rejected, a reference to consciousness cannot be avoided. Given the complexity of this phenomenon, a detailed discussion of its nature would be a topic for another work. In order to clarify Davidson’s and von Wright’s view about the mental nature of mental phenomena and to find out whether it is a plausible view, a few brief remarks about consciousness are, however, mandatory. This is because consciousness seems to be an essential feature of at least some mental phenomena.

Donald asserts: “Georg believes that he is late for the meeting”. In thinking that Georg believes this, one does need not to suppose that he is conscious of the belief that Donald has attributed to him. Georg does not need to be thinking of that belief at the time of the attribution; yet he may nevertheless believe that he is late. It appears that with respect to those beliefs which a person does not believe at time t, it is in order to say that they can be attributed to the person without his awareness of the belief in question. It would not be false to say that a person believes something as long as it is understood that this believing differs from conscious believing. It may be correct to say that Donald believes that Paris is the capital of France, that clouds are made of vapor or that snakes are animals. Donald may believe these things even though he does not believe them at the time of the attribution, or in fact even if he does not ever consciously entertain these beliefs. In saying that Donald believes that “Paris is the capital of France” we do not usually imply that he is conscious of this belief. What we roughly mean is that Donald would answer appropriately if asked the appropriate questions or behave in a way which would agree with the fact that he does have these beliefs according to our standards.

Let us suppose on the other hand, that Georg believes that he is late for the meeting and thinks this silently in his head. This “discussion with oneself” is something which we all find ourselves doing from time to time. This kind of conscious belief is something different than the belief which is attributed to Donald even though he is not aware of the attributed belief. In the case of conscious belief there is a clear subjective element present, which is lacking in the case of non-conscious belief. There is nothing particularly mental about non-conscious beliefs, they can be understood solely as being physical dispositions, each ultimately a mechanism which, given appropriate conditions has, with high probability, such and such consequences. This kind of belief need not be “in the person” at the time of the attribution. On the other hand, in the case of conscious belief there is, supposing that O-
physicalism is true, something physical corresponding to the belief. In this case, there is a physical truth about the mental. However, it seems that a belief or believing of which a person needs not to be aware does not need to be implemented in the brain. Does this matter? It certainly matters when we consider the problem of mental causation. How could a belief which is not physically present in the person, more specifically in the brain, cause anything?

My claim is that a firm distinction between belief and conscious belief, call it belief_C as well as believing and consciously believing, call it believing_C should be made. Beliefs_C are something which are “in” the conscious awareness of a person; they are beliefs which a person consciously believes – whereas beliefs are something that may come to be in the conscious awareness of the person. Of beliefs_C we are immediately aware when we experience or “have” them. The reasons to make these distinctions are that: 1) there certainly are beliefs_C and therefore (so I contend), 2) the only way to make sense of mental causation is by referring to beliefs_C. How should we understand the nature of conscious mental phenomena? I suggest, and this is obviously very crude, that a propositional attitude of which one is conscious is nothing but an internalized sentence that one is thinking of. What does “thinking of” mean? Doesn’t it push the question about the nature of conscious states only a step further? Perhaps, but it is doubtful whether there is any further mental explanation of what “being aware of the existence of a propositional attitude” could mean beyond the observation that it is a form of silent speech. What do we do when we consciously entertain a thought? We think it “in our head”. Is there something more to it? It is difficult to see what more could be said. How would you answer the question “How do you know that you are thinking of X?” Merely by asserting that you are thinking of X. The only criterion that you have for the existence of your conscious thought is the thought itself; there are no other criteria. I believe most of us understand what it means to think a thought silently in one’s head. But at the same time it must be admitted that the nature of introspection is unclear.

What do we do when we introspect our own thoughts, what kind of process takes place when we are conscious of our thoughts? Hacker, who being a Wittgensteinian is very dubious of private experiences, suggests that: “There is such a thing as introspection, but it is not a kind of inner perception – it is a form of self-reflection.” But what is self-reflection? In the

716 The distinction should of course be made with all mental states with propositional content; belief is used here only as an example. As far as sensations are concerned, it seems to me that there can be only sensations_C. A pain of which a person is not aware is an incoherent concept.
717 This claim will be defended in chapter four.
following I will describe von Wright’s brief answer to this question. The answer seems to me congenial.

What are Davidson’s and von Wright’s views with reference to the kind of distinction suggested above? In their writings references to consciousness are few and I suppose that they would see the distinction between belief and belief$_C$ as overtly Cartesian because it assumes that there is an inner truth which is the final arbiter with respect to question of whether a person is really in the state of believing, or whether she really has the belief that is attributed to her. This kind of suspicion follows from their neo-behaviorist views about the mind. Davidson is silent about the nature of consciousness; he mentions the concept only once or twice in his writings. This is quite remarkable given the centrality of the concept in modern philosophy and given the extent of Davidson’s writings. Some critics have seen this as a clear defect of his views and concluded that perhaps Davidson is an eliminativist with respect to qualia and consciousness. It seems to me also that the failure to discuss the nature of conscious states is a flaw in Davidson’s philosophy of mind. If his concept of “mental event” is meant to refer to conscious events, there is no mention of it whatsoever in his writings. What is clear, as we are about to see, is that Davidson’s use of the “mental” does not usually refer to conscious states. This being the case, AM is silent about the phenomena that I would call belief$_C$.

When going through von Wright’s writings for the first time, I had the impression that his position would suffer from the same flaw as Davidson’s, namely an ignorance of and lack of attention to consciousness. However, it turned out that one of the last things that von Wright ever wrote was a paper called “Consciousness”. It was a work which was left unfinished; yet it shows that von Wright’s interests in the philosophy of mind were gradually moving towards the problem of consciousness. I mention these facts in order to emphasize that although von Wright never made this writing public, he nevertheless had carefully articulated views about consciousness. Being aware of this may hopefully preclude possible misinterpretations of von Wright’s position. It is interesting that von Wright makes a somewhat similar distinction to the one suggested above. Especially for those who are interested in von Wright’s work, it is worth noticing that his ideas relating to consciousness supplement his general view about the mind in a way which opens up avenues for further

---

719 For an interesting recent analysis about the nature of self-reflection see, von Kutschera, 2006.
720 See, for example, Robinson, 2001.
721 The paper was found from the archives of von Wright. In the bibliography it is referred as von Wright, UPe.
722 I was surprised when I found out von Wright does make a distinction which I think is essential in order to develop a thorough view about the nature of mental states.
explorations. I think that von Wright’s brief discussion of the nature of conscious states makes his view more interesting than Davidson’s, whose conclusions seem to have counterintuitive consequences. I thus have a more optimistic view about von Wright’s position than what he himself had. The unfinished article was left with a note: “I fear that this paper will stir more misunderstanding than understanding.”

In his last paper, von Wright makes a distinction between consciousness and self-consciousness. The former is something which is a defining feature of living beings, although von Wright acknowledges that it is unclear to which living beings this concept can be meaningfully applied. The criteria for the attribution of consciousness are behavioral. Human behavior and behavior resembling it is the standard against which consciousness-exhibiting behavior must be measured. This is a familiar point from Wittgenstein and its importance has been emphasized by neo-behaviorists like Hacker. Self-consciousness, on the other hand, is a property or an ability which is unique to humans. Von Wright defines self-consciousness as the ability to be conscious of one’s own mental states. Being conscious of one’s own mental state is something like reflecting on one’s mental state. This self-reflecting is essentially linguistic and therefore only possible for humans. The idea of self-consciousness as linguistic reflection is: “…related to the idea of soliloquy or monologue, and to the special form of monologue we call speaking to oneself.”

Von Wright concludes that to be self-conscious is: “[...] like speaking to oneself e.g. saying ‘I have a toothache’ or saying ‘I have not toothache’”, it is like asking oneself: “Do I experience M, or not?” I think the reference to silent speech to oneself must be taken literally and that it is in essential respects a correct description of how the nature of self-conscious states should be understood. In this work I am not willing to elaborate this point further since it would require a thorough analysis of the concept of consciousness. I merely note that I was very interested to see von Wright in the end describing a view which has always seemed congenial to me and which I thought was lacking from his philosophy of mind.

It seems clear that there is a subjective element involved in genuine mental states – that is, in those states of which we are conscious. As von Wright suggests, the subjectivity consists in the fact that any outward evidence that a person is in a state which can be described as belief_C, that a person is self-conscious of a certain thing at time t, is not and

723 Von Wright, UPe, 8.
724 Von Wright, UPe, 15.
725 Von Wright, UPe, 20.
726 Davidson (1970, 211) disagrees: “[...] the distinguishing feature of the mental is not that it is private, subjective or immaterial....”
cannot be conclusive. Only the subject knows what the case is. A non-conscious belief, on the other hand, can be attributed beyond “reasonable doubt”. There are, or can be, reasons which conclusively justify our belief that Georg believes that “Paris is the capital of France”. There is no evidence which would conclusively justify the attribution of the same belief to Georg. To claim that conscious states necessarily have a subjective aspect which cannot be captured from the outside is merely to note that there are different kinds of mental phenomena and, as far as I can see, Davidson’s and von Wright’s conception of mental phenomena fits better with the view that we need not be aware of mental states, although these states can still be attributed to us. This could be taken to mean that consciousness is not an essential feature of mental phenomena. Unlike Davidson’s view, von Wright’s understanding of the nature of the mental captures the essential feature of the mental – its subjectivity. I think we should appraise von Wright’s approach, because I fail to see what reason there could be to call non-conscious phenomena “mental”. Strictly speaking, we should say that there are no such things as non-conscious beliefs just as there are no non-conscious pains. Such “beliefs” deserve not to be called mental. I will return to discussing the importance of this view when the problem of mental causation is addressed in chapter four.

2.5.2 Radical interpretation – Davidson’s conception of the mental

It could be claimed that there is a quite obvious reason why Davidson remains silent about the subjective aspect of mental states. His conception of the mental is based on the view that facts about mental phenomena are publicly available, and thus nothing inner can be of interest in deciding whether a person is in a mental state. Davidson’s famous idea of radical interpretation is based on the conviction that there cannot be facts about what a person means or believes which would transcend our ways of knowing them. The use of a conceptual tool like radical interpretation reflects Davidson’s philosophical interests; he wants to understand how we are able to understand others.

The question of what understanding others involves was a central question in Davidson’s philosophy from the beginning. In the first place, the considerations about interpretation were meant to help to answer a question in the philosophy of language. The question which troubled Davidson during the 1960s was “What is meaning?” Instead of trying to answer this question, Davidson proposed that one should try to give an answer to different question, namely “What would it suffice an interpreter to know in order to understand the
speaker of an alien language and how could he come to know it? Davidson’s description
how a theory of meaning could be verified is expressed in his idea of radical interpretation. In
Davidson’s early writings, radical interpretation was a methodological device meant to help to
clarify questions about meaning. Later the notion of interpretation came to be the central
notion in Davidson’s philosophy. The importance of this concept becomes clear from the
following quote:

It is always possible… to improve one’s understanding of another, by enlarging the database, by adding
another dose of sympathy or imagination, or by learning more about the things the subject knows about.
This is the process of radical interpretation. There is no further court of appeal, no impersonal
objective standard against which to measure our own best judgments of the rational and the true.

The quote shows how interpretation has gotten a central role in Davidson’s view concerning
our understanding of ourselves and the world. In his later works, the importance of
interpretation is highlighted to the extent that it is an indispensable requisite for language and
thought and a necessary condition for acquiring the concepts of truth, objectivity and reality.
Questions about rationality and truth make sense only against the standard which is set by the
intersubjectivity of interpretation. When this principle is applied to the philosophy of mind, it
can be said that questions about the nature of mental states make sense only against this same
standard. What often goes neglected in the discussions about the status of radical
interpretation are the questions which interest Davidson, and of which the approach of radical
interpretation is meant to provide clarification. Davidson notes, for example: “I have long
been puzzled by the question as to what it is about human behavior, verbal and otherwise, that
makes it possible for us to figure out… what others think and mean, want and intend. The
question is what it is about behavior which allows the inference to mental phenomena; a
discussion, for example, about the brain would not answer this question. Davidson thinks that
by discussing the nature of radical interpretation an answer, which goes to the basis of how
we are able to understand other people, can be achieved. As he says:

Radical interpretation will throw light on the question of how we can tell when a creature has a genuine
concept… The point of the study of radical interpretation is to grasp how it is possible for one person to
come to understand the speech and thoughts of another, for this ability is basic to our sense of a world
independent of ourselves, and hence to the possibility of thought itself.

---

727 Davidson, 1994, 126.
728 Davidson, 1994a, 232, emphasis mine.
729 Davidson, 1999b, 155.
730 Davidson, 2001f, 123, 127. Compare: “I want to know what it is about propositional thought that makes it
intelligible to others.” (Davidson, 1995a, 14)
I would claim that these questions are of fundamental interest to a radical interpreter. The main motivation behind radical interpretation is not, as for example Glock claims, the

\[\text{[\ldots] hope of accommodating apparently recalcitrant phenomena like meaning and thought within a naturalistic framework. If radical...interpretation is possible even in principle, this will show how our rich intensional and semantic concepts and statements can be derived from, albeit not strictly reduced to, something more basic.}\]^{731}

Glock continues: “If we can show that meaning and understanding are possible on an austere extensional and physicalist basis, the threat which these higher level phenomena pose to a naturalistic world-view seems to be defused.”^{732} In his study of Quine and Davidson, Glock writes that Davidson regards as problematic concepts that are both semantic and intensional or linguistic (like meaning, synonymy and reference). These concepts do not appear on Davidson’s list of elementary concepts. Given that the intentional supervenes on the non-intentional, the purpose is to bridge the gap between them. This would mean that the puzzling phenomena which exhibit intentionality could be analyzed in terms of more basic, non-intentional phenomena. A theory about problematic phenomena like thought should be based on more elementary evidence. Ultimately this evidence is the physical behavior of people, which can be interpreted without mental concepts. If this willingness to promote extensional concepts is indeed present in Davidson, it is plausibly interpreted as a remaining piece of influence from Quine and the positivists. Indeed, Glock claims that Davidson’s alleged move away from Quine’s behaviorism is “half-hearted and does not make interpretation from scratch intelligible.”^{733} Again, this interpretation is similar to the one suggested by Mullhall, who claims the following of Davidson’s view:

The world we really perceive is radically devoid of any human significance, until we use our interpretative theorizing to organize…primitive data into units of human meaning – words, actions,

---

731 Glock, 2003, 169, my emphasis.
732 Glock, 2003, 249, my emphasis. This view sounds similar to Mullhall’s (1987, 322): “Davidson’s commitment to the notion of bare sounds and bare movements is strikingly analogous to empiricist sense-datum theories of knowledge: in both, it is presupposed that everyday experience of the world can be illuminatingly viewed as a logical or theoretical construction out of brute data – “the given”; and in both… an unsubstantiated belief in the possibility of a systematic redescriptions of the everyday world in terms of this postulated category of brute data is held to reveal something fundamental about the ontology of that world…. If this notion of the given is juxtaposed with the picture of alienation from language and other people, we can see how much of Davidson’s world-view is the result of Quinean influence.” Glock thinks that, according to Davidson, intensional concepts pose some kind of threat to a “naturalistic framework” which can however be solved. Mullhall makes the same point when he notes that Quine and Davidson want to show that intensional concepts are no threat to a physicalistic worldview. Indeed, by showing how intensionality reduces to “brute data” we would show something fundamental about the ontology of the world.
733 Glock, 2003, 188. On the other hand, puzzlingly enough, Glock (2003a, 353) admits that “the perspective of the radical interpreter is not unduly behaviourist or verificationist.”
gestures...the significance and the humanity we find in...phenomena of our everyday life are a result of our reading our concepts into the data we directly apprehend.  

I think we should strongly disagree with this kind of interpretation. Radical interpretation does not provide, and is not meant to provide, a reduction of intensional concepts to something else although, in Davidson’s words, it is “an important step in the direction of reducing complex and relatively theoretical intensional concepts to intensional concepts that in application are closer to publicly observable behavior.” Notice: intensional and closer, but still not reducible, to publicly observable behavior. Against Glock, I must doubt that Davidson would have thought that higher level phenomena pose a “threat” to a naturalistic world view. In order for them to be a threat, one would first have to accept the idea that lower level phenomena are somehow more natural than others. Davidson, however, stands firmly against this view. There are no epistemological priorities between the descriptive levels, and we are equally in touch with reality through different ways of describing it. I admit that this interpretation is somewhat controversial because, as we have seen, in Davidson’s philosophy of mind there is a certain bias towards the physical.

It should also be recognized that Davison claims that it would be desirable to base semantic theory on non-semantic evidence, because this kind of “reduction” would make for “conceptual progress.” In the same context Davidson notes, rather cryptically: “An attempt to build [the theory] on even more elementary evidence, say behavioristic evidence, could only make the task of theory construction harder, though it might make it more satisfying.” I do not want to speculate on the question of how this remark should be interpreted, but I admit that it could support the view that Davidson is a true Quinean-behaviorist or eliminativist in disguise. However, this would go so clearly against the spirit of Davidson’s overall anti-reductionism that I reject the possibility.

---

734 Mullhall, 1987, 322.
735 Davidson (2001f, 132) notes: “Of course, propositional attitudes are involved; they just aren’t expressed, in the theory, in a way that individuates attitudes generally, and in a way that would make the theory circular.”
736 Davidson, 1980, 4, my emphasis.
737 Davidson (2001d, 11) notes: “Our failure to provide an analysis of the concept…of thought…does not mean there is something hopelessly mysterious about [this concept]; it only reflects the fact that intentional phenomena cannot be reduced to something simpler or different.”
738 With respect to the purpose of radical interpretation, commentators have drawn interestingly opposite conclusions. Glock (2003, 182) claims that: “Davidson seeks to extract rich semantic concepts and statements from evidence which he considers to be more basic, because it can be described in non-semantic terms.” This should be compared to Evnine (1991, 101): By admitting that the evidence for a theory of meaning is not entirely non-semantic, Davidson is opposing reductionism. Semantic facts cannot after all be extracted from purely non-semantic facts, as behaviorists and other reductionists have hoped.”
739 This problem will be considered in section 2.7
740 Davidson, 1974a, 142.
741 Ibid.
As far as the details of radical interpretation are concerned, the original idea of radical interpretation is wholly theoretical and methodological. Its purpose is to give a suggestion of how one could come to know that an interpretee means one thing or another, or has one belief or another. A philosopher is looking for a fundamental answer and this kind of answer cannot be had if too many of the things that she wants to explain are assumed. Only by taking the position of radical interpreter can we hope to answer the question of what the conditions on the possibility of understanding are. What is relevant to the question about the nature of mental states is that one cannot assume that one knows the content of those states before interpretation has started. One does not start with a class of mental states as though it were already transparent what those states are. Instead, if one is looking for a fundamental answer, one has to start without any assumptions. This remark can be applied to the case of mental properties and to the problems they allegedly pose for non-reductive physicalists. Kim and others start with the assumption that the nature of mental properties is transparent and then go on to ask what their relation to physical properties is. However, I have suggested that this approach errs in its assumption that we understand clearly the nature of mental properties.

The fact that the available evidence for the attribution of mental phenomena must be restricted led Davidson to invent the thought experiment of radical interpretation. Originally, the problem was formulated in terms of the question of how a field-linguist could come to understand a speaker of a foreign language. The motivation behind this thought-experiment should be clear. One wants to try to find out what it is about thoughts and meanings that makes them interpretable and understandable. What is it about humans and their behavior that enables us to know what other people think and mean? People are not the kind of mind-readers described in science fiction stories and they cannot assume, at least not when doing philosophy, that they know what other people are thinking. If a conclusion that people in fact make such assumptions is reached, a philosopher wants to know on what these assumptions are based. This, incidentally, is a good example of Davidson’s philosophical strategy in general. He wants to know how understanding is possible, in principle. Davidson is not suggesting that radical interpretation models the actual competence of a field linguist, child, or an adult interpreting the linguistic behavior of another with a homophonic translation manual. He says: “The approach to the problems of meaning, belief and desire that I have outlined is not… meant to throw any direct light on how we master our first concepts and our first language”.\(^\text{742}\) However, it should also be noticed that all the three cases mentioned above

\(^{742}\) Davidson, 1980, 12. Exactly the same point is repeated ten years later (Davidson, 1990, 325), so it is clear that on this issue Davidson’s views remained steady. He repeatedly emphasizes that the model of radical
are instances of radical interpretation. The case of a field linguist, made famous by Quine, should be straightforward. A non-linguistic child is also an example of a radical interpreter, since the child has to go on without any prior knowledge of the meanings or beliefs of the people he is observing. Finally, according to Davidson, an everyday interpretative situation is also an instance of radical interpretation. Doesn’t this sound contradictory? On the one hand it is being repeatedly claimed that radical interpretation is not meant to throw any direct light on how we actually understand other people; on the other hand, Davidson claims that all understanding involves radical interpretation.

I believe there is a simple solution to this confusion. Davidson has noted that there is an “official” and “unofficial” story about radical interpretation. The official version is the theoretical idea that Davidson first suggested in his article “Radical Interpretation”. According to this official formulation, it is essential to show how a theory of meaning can be interpreted without appeal to any evidence that assumes the individuation of the contents of any propositional attitude. The merit of the official story lies not in its plausibility as an account of how a person actually understands others, but in the fact that it amounts to an “informal proof” of the adequacy of the theory to yield what is needed to support the interpretation of basic propositional attitudes. This is an idealized model, and it does not attempt to say anything about how people actually operate in actual situations of interpretation. The purpose of the model is to describe sufficient conditions for understanding.

The “unofficial story” is closer to our intuitions concerning the details of everyday understanding. People never have enough of the sort of evidence which would be needed to follow the official route, and they always have a great deal of other sorts of evidence. The claim that all understanding involves radical interpretation is meant to highlight only that what someone means is always an empirical question, since there are no such entities as determinate meanings. There is no more in meaning than what can be found out in the process of interpretation. Understanding requires interpretation, and a theory of interpretation is a general theory of what understanding people involves. Davidson has detailed views about the question of what understanding requires. They need not concern us here.
The famous slogan that “meaning is use” has been interpreted in many ways. However, to acknowledge that use creates meaning explains the focus on radical interpretation. The claim “meaning is use” is not a definition of meaning; it is rather an acknowledgement of the view that attributions of meaning must be tied to observable behavior. It is to acknowledge that there are no mysterious entities ‘out there’ called meanings. Words do not have meanings intrinsically, they have meanings only because they are used in certain ways. What holds for meaning, holds also for belief. Davidson argues that behavior is the main evidential basis for attributions of belief, and that the attributions of beliefs are supervenient on behavior. This does not amount to traditional behaviorism. Davidson notes: “Propositional attitudes can be discovered by an observer who witnesses nothing but behavior without the attitudes being in any way reducible to behavior.” Although attitudes are not strictly reducible to behavior, there are “conceptual ties”, which are sufficient to allow inferences from behavior to the attitudes. A complex enough pattern of behavior, or a reason to believe that there could be such a pattern, is required for the attribution of a single thought. A strong claim is made when Davidson argues that “unless there is actually such a complex pattern of behavior, there is no thought.” Three theses of neo-behaviorism influence Davidson’s views. First, mental states supervene on behavior. Second, because of “conceptual ties”, behavior is sufficient for inferences to attitudes, and, finally, there must be a complex pattern of behavior found if mental states are to be attributed to somebody or something. Davidson does not object to a mild form of behaviorism: “Behaviorism is objectionable only if it maintains that mental states are nothing but the phenomena we normally take to be evidence for them; or that mental concepts can be explicitly defined in terms of the behavioral concepts.” Davidson is not committed to these doctrines and claims that accusations of behaviorism aimed at him are therefore misguided.

What is the main reason to insist that evidence for the existence of mental states must be publicly available, at least in principle? It is the conviction that language is “intrinsically

[footnotes]

746 Davidson has the view that there cannot be thought without language. The reasons for this are too complicated to explicate here. Let us therefore only note that the connection between language and thought, according to Davidson, is essential. Also von Wright thinks that thought requires language, although he does not give as clear reasons for this view as Davidson. For Davidson’s reasons see for example Davidson 1975, 1990b, 1991, 1992, 1994a, 2000, and the views put forward in interviews with Gluer, 1995 and Borradori, 1994.

747 This raises an interesting question about the exact nature of Davidson’s supervenience, which will be considered in section 4.1.1.

748 Davidson, 1982, 100.

749 This reminds us of Wittgenstein (1992, 63e): “The inner is tied up with the outer not only empirically, but also logically”.

750 Davidson, 1982, 100.

751 Davidson, 1986, 200, my emphasis. Perhaps the same could be said about “nothing but” materialism.

752 This criticism against Davidson is repeatedly raised by Fodor.
social.” Davidson claims that “everyone” can agree to the following: “Whatever there is of interest to the notion of linguistic meaning has to be something that one person can figure out about another, because the whole point of language and meaning is communication.” Since language is essentially a tool for communication the claim is that: “whatever it is about speech that makes it useful for communication, has to be conveyed by the publicly observed behavior of language users.” Given the view that communication is the main function of language, it is only a short step to claim that: “Language is a social art which we all acquire on the evidence solely of other people’s overt behavior under publicly recognizable circumstances,” or that “public availability is a constitutive aspect of language,” or that “language is necessarily a social affair.” The consequence of this is the view that meaning has its life only in successful cases of communication. There is no “real” meaning which could completely transcend the powers of recognition. For a child learning a language, the only evidence available comes in the form of observable facts. Usually from this starting point the child eventually becomes an adult capable of using language fluently. We learn our language by observing the linguistic and non-linguistic behaviour of others. Davidson notes: “[…] the irreducible element of behaviourism implicit in the radical interpreter’s point of view is that what can be learned of how others react to the world and use words is learned by observation of what those others do.” This claim would be challenged by “nativists” like Chomsky and Fodor; here I am not taking a stance on the question but merely describing what Davidson’s view is. The emphasis on behavior means that non-observable facts cannot add any relevant information to linguistic meaning. The facts that are not observable in overt circumstances are irrelevant to linguistic meaning, at least as long as the concept of meaning is not turned into a technical notion that is separated from the actual use of language. Because meaning and thought are social phenomena, the evidence for them must be publicly available. The public availability should be taken literally; the evidence must be available to anyone who is capable of understanding another speaker. This explains why brain states, for example, are not really publicly available, although they are available in principle.

754 Davidson, 2003, 284-285, my emphasis.
756 Davidson, 1990, 314. Davidson (ibid.) refers especially to Wittgenstein as the source of this view: “As Ludwig Wittgenstein and many others have insisted, language is intrinsically social. This does not entail that truth and meaning can be defined in terms of observable behavior, or that it is ‘nothing but’ observable behavior; but it does imply that meaning is entirely determined by observable behavior.”
757 Davidson, 1992, 117.
758 Davidson, 2003a, 692.
759 Davidson (2000) himself notes that those who embrace conceptual atomism may merely have a different understanding of the question what counts as a “concept”.

265
An approach to the mind which emphasizes the necessity of interpretation has important consequences for the question of what the nature of mental states is. The third person perspective on mental phenomena emphasizes the inevitable intersubjective and objective elements of mental phenomena. In this approach, there is perhaps no place for the subjective awareness which I have suggested as being an essential feature of mental states. Davidson writes: “[…] what a fully informed interpreter could learn about what a speaker means is all there is to learn… the same goes for what the speaker believes.”

What a fully informed interpreter could learn captures the whole truth about the beliefs of an interprettee. This could be seen as an unjustified move from an epistemological premise to an ontological conclusion. Colin McGinn has raised this criticism against Wittgenstein and Antony claims that non-reductive physicalists are confusing epistemological and ontological issues.

Here Davidson’s understanding of “intentional realism” creates a tension between him and those naturalists who claim that there must be interpretation-independent facts about mental phenomena. It seems to me that if we consider how psychological language is usually used or consider how language in general seems to be learned, Davidson’s realism agrees with our intuitions better than the view claiming that the facts about mental phenomena must be interpretation-independent. Wittgenstein once wrote that the essence of the mental is not something that can be clearly shown; only its different features can be described. He also suggested that by investigating the laws of evidence for the mental we are investigating the essence of the mental. The way I read these remarks is that there is no “essence” of the mental; the only nature of the mental that we can find is a result of our describing certain important features of mental phenomena. What a Davidsonian approach is attempting to do is to describe these features. Davidson claims: “[…] speakers of natural language can be, and often are, correctly understood on the basis of non-linguistic facts not merely available, but readily available, to the likes of you and me.” This kind of view is justified if we consider how linguistic practices actually work. Also here Wittgenstein’s lesson seems relevant. He suggested that instead of guessing how a word functions one should look and see how a word or expression actually works in the life of humans. Instead of speculating, we should consider: “is the word ever actually used in this way in the language-game which is its original home?”

Davidson argues that one can be a realist about mental phenomena who thinks that

---

763 Davidson, 1994, 126.
there are correct and incorrect interpretations of what a speaker means by his words and what he thinks. This version of realism is based on the conviction that the primary and ultimate source of meaning lies in successful interpersonal communication. The original home of the language game of the mental is the language of the everyday, and a realistic conception of the mental can be built on this ground. This kind of realism looks suspicious only if one has a “Cartesian, individualistic conception of meaning and the intentional….” I think Davidson’s realism seems inappropriate only if we think that what someone means or believes is absolutely independent of what is understood by others, or that expressions have meanings that are independent of the facts about how people understand each other in the course of life. According to this view, it could turn out that the evidence for the attribution of meanings and beliefs could always lead one astray. But, so I think, we must stand on some ground; we manage to communicate most of the time and we successfully attribute mental phenomena to others. This obvious fact can be our starting point.

It may be difficult to justify going from an epistemological premise to an ontological conclusion, but it is difficult to challenge the view on which the conclusion is based – namely, that the best interpretation an interpreter can produce is as objectively correct as it can be. The objectivity of interpretations could be challenged if we had another view of what beliefs are “really like”. The objectivity of attributions based on the folk-psychological method could be challenged if we had a convincing alternative which describes what mental phenomena really are. Eliminative materialists, for example, claim that they do have a conception of what beliefs really are: they are mere fictions. As a challenge to this suggestion, we should ask on which conception, on the folk-psychological one or on the eliminativist one, are our lives and practices actually based?

2.5.3 Interpretation and the nature of mental states

Davidson’s position has been perspicuously termed as a form of “third-person Cartesianism”, the basic idea being that under ideal conditions, people can have complete and infallible access to the mental lives of others. Davidson writes: “Thoughts, desires, and other attitudes are in their nature states we are equipped to interpret; what we could not interpret is

---

765 Davidson, 1994, 127.
766 This expression is from Byrne, 1994, who uses it to describe a contemporary form of behaviorism which he attributes especially to Davidson and Dennett.
not thought.” This is a strong claim which expresses Davidson’s understanding of the concept of thought or of the mental more generally. Criticism against this understanding can obviously be raised and Davidson admits that: “I (perhaps somewhat arbitrarily) restrict the word ‘thought’ to mental states and events with propositional contents.” Davidson is thus trying to define the concept of thought. Davidson’s understanding of how concepts can be defined was described in section 1.2.2.

If we accept that a third-person view of mental phenomena is mandatory, what can be said about the nature of mental states that are being attributed as a result of interpretation? From considerations of how interpretation has to proceed, Davidson draws various conclusions. Interpretation must proceed so that the interpreter uses his own norms in trying to understand a person. A caveat with respect to the expression how “interpretation must proceed” needs to be noted. As I showed in the previous section, at times Davidson claims that his proposal should be understood as a suggestion about how interpretation could proceed, and it is not meant to describe how humans actually master their first concepts or learn language. In this sense, it could be argued that Davidson is not claiming how interpretation must proceed; he is not providing necessary conditions for interpretation. But it is also clear that various suggestions of Davidson’s show that for him the approach of radical interpretation is a mandatory perspective. The third person approach to language and thought is not a mere philosophical exercise.

An interpreter has to consider how to make best sense of the person to whom mental states are being attributed. In the act of interpretation, the causes of beliefs have to be evaluated as well as the logical or conceptual relations between beliefs. These relations specify the content of a specific belief. A thorough discussion of holism, externalism, and normative constraints would be required for a full picture of the Davidsonian conception of the mental. However, such discussion is not essential in order to understand what the general nature of mental phenomena and their relation to physical phenomena are. A clarification of this issue is essential for a discussion of the irreducibility of the mental and the problem of mental causation.

---

767 Davidson, 1990b, 14.
768 Davidson, 1999k, 327.
769 For Davidson’s definition of thought see Davidson, 1989, 1990b and especially, 1995.
770 On the other hand Davidson (1994b) notes that by not describing actual practices he tried to find out and clarify what is necessary for linguistic communication.
771 I explored these questions in my Master’s thesis (Kuusela, 2001).
772 How externalism, normativity, and holism relate to the irreducibility of the mental will be discussed in chapter three.
The only measuring stick that can be used when interpreting others are our own beliefs. We understand others if we can make some sense of them from our point of view. This is not the case when we try to understand or explain physical phenomena. The process of attributing mental states to others is called interpretation for a reason; two systems of beliefs need to be correlated. The nature of the physical can be studied by making observations, whereas the interpetee’s belief system is understood only to the extent that it makes sense from the perspective of the belief-system of the interpreter. The two belief-systems must “fit” somehow; if we could not understand the other at all, what would be the reason to attribute mental states to him (it?) at all? Truths about the physical can be, and in fact have to be, established without interpretation. These truths are a result of observing how things are. In the case of mental phenomena, there are no such interpretation-independent truths because facts about the mental are construed in the process of interpretation. What linguistic or non-linguistic behavior means is always an open question, for which multiple answers can be given. Subjective decisions must be made when interpreting others. The use of subjective evaluation is not necessarily required in the course of understanding others in the everyday life, but it remains as a principled difference between understanding others and explaining mindless nature. Why think that this is how the nature of the mental should be understood? I believe that most of us are familiar with the concept of ‘thought’; we have an intuitive grasp of this concept. Given what we mean by it, Davidson’s conclusion is that: “A creature that cannot in principle be understood in terms of our own beliefs and mode of communication is not a creature that may have thoughts radically different from our own: it is a creature without what we mean by thoughts.”

Can this view be shown to be true? Perhaps not, but if the Davidsonian conception is rejected, then the challenge is to explicate what we do mean by our concept “thought”. Why use this term to describe states that we cannot understand? This is partly a terminological question, but it also challenges us to contemplate the nature of our actual concepts.

Many critics have argued that emphasizing the primacy of interpretation has the consequence that mental facts turn out to be less real than other facts. Antony has claimed that from the perspective of AM, a mental–physical identity cannot be a “genuine fact.” The question which directly concerns the reality of mental phenomena is this. Does interpretation create order in the belief system of the interpreter, an order which did not exist before interpretation started? I think this could be argued. Davidson namely notes: “It is an artifact
of the interpreter’s correct interpretation of a person’s speech and attitudes that there is a large
degree of truth and consistency in the thought and speech of an agent.”

Consistency of thought is a prerequisite of interpretation, because without some consistency no sense of the
other could be made. On the other hand, Davidson writes: “Sentences have logical relations
with other sentences, and interpretation must, as far as possible, preserve these relations.”

My question is whether there is (already) consistency in the thought and speech of the other
person and we preserve it in interpretation, or is the consistency an artifact of interpretation?
If the consistency of interpretee’s thought is a result of the interpretation-process, doesn’t this
imply that the reality of mental states is of a different sort than the reality of purely physical
states? If events are mental only as described, is it correct to say that the existence of mental
events is language-dependent whereas the existence of physical events is not?

Given the “constructivist” element of interpretation, it seems unclear what Davidson’s
position with respect to the reality of mental states is. The question of Davidson’s relation to
intentional realism has been an issue under debate, dividing critics considerably. It is
interesting that the conclusions are so diverse; I believe the reason for this is the difference of
opinion as to what “intentional realism” requires. Many, for example Antony, Klagge, Rawling,
Jacob, Skillen, and Campbell see Davidson as a mental antirealist. This form of antirealism could be associated with a form of eliminativism. What these critics share is the view that the nature of interpretation, especially the requirement that interpretation
should proceed according to normative standards, leads to antirealist results. Others, like

---

775 Davidson, 1983, 150, my emphasis.
776 Davidson, 1999, 307, my emphasis.
777 For this question see Emiliani, 2001 and Up.
778 Antony (1989, 179) claims that, according to Davidson, “one should look upon psychological hypotheses not
– as one views hypotheses in the physical sciences – as our best guesses about the nature of an objective
phenomenon, but rather as artifacts reflective of our current stage of self-interpretation.” Antony concludes that
given the “radically constructivist” position of Davidson, Dennett’s theory of mind is what Davidson should
really have. We should note that Davidson (1997c) explicitly distanced his position from that of Dennett’s.
779 “We ascribe beliefs and desires to people, in part, as a way of understanding, predicting, and appraising their
behavior. Thus, the mental becomes more a way of seeing people than it is something in people that can be
seen… this aspect of Davidson’s views pulls away from his seemingly realistic conception of the mental”
(Klagge, 1990, 342).
781 Jacob, 1997.
783 “Since an individual’s mental states depend on an interpretation of that individual’s behaviour… and the
states ascribed at any one time may later be over-rulled if the interpretation requires alteration in light of new
behaviour… there is no fact of the matter about someone’s mental states. This dependence of mental states’
ascription on interpretation is not unlike a form of anti-realism since on this view mental states become
recognition – (or interpretation-) dependent.” (Campbell, 1997, 379) Campbell concludes that it is unlikely that
one could find “compelling reasons” to believe that Davidson is a realist about mental properties. In making this
claim, Campbell is one of the commentators who correctly see that the antirealism could be seen as a result of
Davidson’s general attitude towards properties.
Joseph, Dennett, and Heil interpret Davidson as a realist. They think that this position follows from Davidson’s monism. If each mental event is identical to some physical event, how could the former be less real than the latter? Davidson himself has raised this question by asking: “How could there be a question about the ontology of mental entities for me if, as I hold, mental entities are identical with entities we also describe and explain, in different terms, in the natural sciences?” We should note how Davidson here ties ontology to natural sciences. Some e.g. Kim, Evnine, Hornsby, and Melchert see Davidson’s position as a middle position between antirealism and realism. They claim that Davidson’s theory of mind is not easily labeled as realist or antirealist. With this interpretation I am essentially in agreement. As I mentioned in section 2.5.2, Davidson’s form of intentional realism is of a peculiar sort.

But why do so many critics see Davidson as an intentional antirealist? I think that the essential question that must be considered is that of what realism with respect to mentality requires. Critics seem to conclude that interpretation-dependence leads to irrealism because what “truly exists”, what is real, must exist independently of us. In my view, the question of how the reality of mentality should be understood in the Davidsonian context is very well captured by Melchert:

What then is an intention or a thought? It is whatever we correctly call an intention or thought. But what is it really? Here is the anomalous monist’s response to that question; if by ‘really’ you mean “apart from all descriptions”, there is not and cannot be any answer; but if you mean, What is the really correct description of it? You must specify the regulative and constitutive principles for some system of description, and then it will really be what the system describes it to be. So a certain event is really a thought (relative to the intentional principles of description), and it is really a brain process (relative to physical principles).
I believe this description captures the Davidsonian spirit, which is somewhat reminiscent of Wittgenstein who in the Blue Book entertains the following thought experiment.\textsuperscript{793} Let us imagine that a person could observe her own brain while entertaining a thought. She sees the thought by seeing her brain and at the same time she is having the thought, that is, thinking silently something like “I have a toothache”, to use von Wright’s example. Wittgenstein asks, in this kind of case: “[…] is the subject… observing one thing or two things?”\textsuperscript{794} He notes that one of these things the subject could perhaps call “a thought”, the other experience would be seeing her brain work. Wittgenstein claims: “Both these phenomena could correctly be called ‘expressions of thought’; and the question ‘where is the thought itself?’ had better, in order to prevent confusion, be rejected as nonsensical.”\textsuperscript{795} Von Wright, when discussing a similar kind of thought experiment, agrees: “The ‘identity thesis’ now maintains that your acoustic sensation is identical with your percept (identical with that which you see in the brain mirror). How can this be anything but a tremendous confusion!!!”\textsuperscript{796} Where the thought really is, can be answered only if “where” is first interpreted. To ask, outside all descriptions, what or in the case of thought, “where”, something really is, is a confusion because the answer will always be one or the other of the descriptions. Descriptions in mental terms are in no way less real than descriptions given in physical language. To think that the latter would be somehow more real is to already commit oneself to a certain understanding of what realism requires or what being real means. It is to accept a certain picture which, when accepted, constrains philosophical investigations. Thought is no more physical than mental, it is not “really” anything, although it can be described in various different ways. Melchert’s description of anomalous monist’s response to the question of “what a thought really is” resembles Rorty’s more general analysis of Davidson’s position. According to Rorty, the line of argument in “Mental Events” leads to the following conclusion:

\[\ldots\] reality does not have an intrinsic character, but can be described in any way… that language-users find useful. None of these ways is more faithful to what is described than any other, nor are there philosophical problems about how these various descriptions mesh.\textsuperscript{797}

Whereas I agree that this is how an anomalous monist could respond when charged with the “what thought really is” – question, it is not entirely clear that Davidson’s physicalistic

\textsuperscript{793} Wittgenstein, 1958.
\textsuperscript{794} Wittgenstein, 1958, 8.
\textsuperscript{795} Ibid.
\textsuperscript{796} Von Wright, UPg, 25.
\textsuperscript{797} Rorty, 1998, 390.
ontology allows him this maneuver. I would also like to emphasize that there is a kind of antirealism in Davidson’s view about the mental given that “the process of specifying the content of a thought... does not require that we suppose there is a definite... object before the mind of the thinker....” It seems that on Davidson’s account the idea of a specific content, which would be present for the subject, his “private property”, needs to be rejected. One could argue that this makes the content of the belief less real in the sense that there is no determinate and final answer to the question of what the content is.

I would claim that we are familiar with the idea of a specific content through introspection, or self-reflection as it could be called. My thoughts do have a determinate content. I know what I think when I am having beliefs$_C$. When I am consciously thinking something, I know what it is that I think, and the content of that thought is precise. Davidson would say that this experience of subjectivity is due to the fact that I must interpret the thoughts of others, while it makes no sense to say that attributing thoughts to myself involves a process called interpretation. Von Wright would agree with this claim. But if this is true, then there is an asymmetry with respect to beliefs and beliefs$_C$ – and even with beliefs$_C$ – when seen from the first-person and third-person perspectives. The asymmetry has something to do with the fact that I am a sensing subject and I have a special access to my own mental states, which the outside observer lacks. When I have beliefs$_C$, the existence of which do not depend on interpretation, I cannot be mistaken about the fact that I have a specific belief$_C$. This is because having a specific belief$_C$ is nothing other than having in mind a sentence which expresses the content of that belief. There is therefore no way that one could be mistaken about the content of belief$_C$. But it is, of course, possible that interpretation does not capture that specific content and yet, pace Davidson, if the interpretee has a clear sentence in his mind we cannot insist that belief$_C$ does not have a specific content. For the purpose of understanding others, the fact that the specific content cannot be captured from the outside is obviously not fatal. People can make sense of each other while rejecting the idea that the contents of two minds have to match perfectly. Indeed, Davidson claims: “Success in interpretation is always a matter of degree; the resources of thought or expression available to an interpreter can never perfectly match the resources of the interpreted”.

---

798 Davidson, 1995, 216.
799 See von Wright, 1994, 144.
800 Davidson, 1994a, 232.
The thing to bear in mind is that interpretation... the understanding of others is a matter of finding a best fit. As interpreters, we have our sentences and we assign them... to the sentences and attitudes of others, thereby giving those sentences the content we deem best captures what they mean and think. We won’t understand them unless we find a core of agreement, but agreement isn’t the aim. The aim is understanding... there is no reason to suppose the understanding of another’s utterances is ever prefect, nor does it need to be. The myth of meanings summons up the dream of perfect understanding which nothing in the actual process of conversation justifies.  

For the sake of the argument, let us agree that this is a plausible view. According to an approach which emphasizes the inevitability of interpretation, we interpret people as having thoughts and other propositional attitudes – in terms of which we explain their behavior. As I showed in section 2.4.3, Davidson claims that beliefs are not little entities lodged inside the brain. An important reason for this claim is that in interpreting others we need not think of mental phenomena as entities; we should rather see these phenomena as states of whole person. William Child, who labels the interpretation-based approach in the philosophy of mind interpretationism, notes: “[...] there is no license for thinking of... belief and desire as internal states or entities.... For, according to interpretationism, propositional attitudes are not entities or items at all; and if we think of them as states, we think of them as states of people....” The reason for claiming that beliefs are not in the brain is thus that the process of interpretation in terms of which beliefs are attributed does not imply that beliefs would be entities. Because it is also claimed that interpretation captures the whole truth about the mental, beliefs are what a correct interpretation concludes them to be, and there is no justification for the view that beliefs could be thought of as internal entities. The mainstream view in the contemporary philosophy of mind is of course that beliefs are states of the brain. But, if we follow interpretationism, this view becomes suspect. As Child notes:

[…] by focusing on the idea of belief as a property of a whole person we do away with the conception of beliefs as internal states, states which are literally inside people’s heads.... Interpretationism... stands opposed to the view of propositional attitudes as internal states.

An important consequence, again noted by Child, is that: “The attitudes we cite in explaining an action need not correspond to anything running through the agent’s head before she acted.” But, it seems to me that if this is the case then interpretation fails, at least from time to time, to attribute beliefs to a person because these beliefs do run through an agent’s head before she acts. However, we should note that according to Child, the attitudes to which an

---

802 Child, 1996, 123.
804 Child, 1996, 123.
interpretationist refers in his explanations of others’ actions need not correspond to anything in the agent’s head. This, presumably, does not exclude the possibility that the attitudes may exist in the head of the interpretee. Yet, what Davidson himself says about the mode of being of mental phenomena certainly gives the impression that these phenomena apply to the whole person, and it does not make sense to think of them as parts of the brain. How seriously can this suggestion be taken if O-physicalism is accepted?

2.5.4 Consequence of interpretationism: Mental phenomena are not states of the brain

My claim is that the reason why the mental mode of being of mental phenomena is important is that if the consequences of interpretationism are taken seriously, no clear sense of the distinction between beliefs and beliefs can be made. If attitudes do not correspond to anything going on in the person’s head, then their reality can be questioned, as many critics have indeed done. It seems to me that there is an internal conflict in Davidson’s views; on the one hand he has a strongly realistic view about mental phenomena, while on the other hand his interpretationism seems to pull to an antirealist direction. Davidson claims that he is a realist about psychological descriptions, and that facts about mental phenomena are as objective as facts about anything else. Yet, it is also being claimed, for example, that:

>[…] it is natural to think of radical interpretation as taking for data not attitudes but the observed ways the agent changes (acts) as the environment changes…Given such data…we then hypothesize a set of attitudes to explain what we have observed, attitudes which will help predict how people will act in the future.  

Davidson also notes: “We, watching a creature adjust its behavior to its needs and opportunities, read into that behavior the beliefs and desires and intentions that naturally occur in animals with thoughts.” These views sound like instrumentalism. If the set of attitudes is hypothesized, there remains a doubt about their existence. If it is not required that the agent be aware of the attitudes which are being attributed to him, it is always possible that the attitudes are non-existent. If they are non-existent, how can they play any role in the

---

805 Davidson, 1999, 654, my emphasis.
806 Davidson, 1999c, 209.
causing of behavior? I believe we have reasons to think that mental states are causally related internal entities.\textsuperscript{807} In its denial of this, interpretationism errs.

The conflict of interpretationism and monism is also evident. If, according to interpretationism, mental phenomena are states of the person, what is the reason to claim that mental events are also neurophysiological events? It is very difficult to get a clear view from Davidson or von Wright which would help us to determine what kind of entities mental phenomena are. We are told, by Davidson, that: “Having a belief is not like having a favorite cat, it is being in a state…”\textsuperscript{808} Von Wright has essentially the same view: “Having a sensation is not like having an ‘object’ – a hat, say. Perhaps the best answer to the question what having a sensation is, is to say that it is a state in which a sentient being is.”\textsuperscript{809} When we describe a person having a belief, we describe him as being in a certain kind of state. What can be said of this state? Davidson explains: “To have a belief is to have a certain property. From an ontological point of view, for a person to have a belief is just what it is like for a room to have a temperature.”\textsuperscript{810} The point of the temperature analogy is to emphasize the following. When talking about the temperature of the room one does not need to suppose that there is an entity which is the temperature. Likewise, in thinking and talking about the beliefs of people we do not need to suppose that there are such entities as beliefs. Critics, e.g. Rawlings\textsuperscript{811} and Brandl\textsuperscript{812} have questioned, correctly I think, whether this view implies a kind of antirealism with respect to beliefs. One could of course simply go along with Davidson when he claims that beliefs are states and states are not entities. But this kind of talk does not explain much, because we are lacking a detailed description of the nature of “states”, “events”, “entities”, and “objects”. It is unfortunate that Davidson uses these expressions loosely. In my opinion, the analogy between temperature and belief is not very illuminating. What does it mean to say that to have a belief is just what it is like for a room to have a temperature? It could be argued that there are essential differences between the properties of having a belief and having a temperature. The analogy is especially obscure because it is made from an “ontological point of view”, the nature of which Davidson does not further elaborate.

A belief is a state. A mental event, on the other hand, is a change in a mental property. But to say that people have mental properties is to say that certain psychological

\textsuperscript{807} The specific reasons for this will be given in chapter four. It should be noted, however, that my claim applies only to those mental phenomena of which we are conscious. Incidentally, if all genuine phenomena are conscious, it means that all mental phenomena should be understood as causally related internal entities.

\textsuperscript{808} Davidson, 1997c, 74.

\textsuperscript{809} Von Wright, 1998, 50.

\textsuperscript{810} Davidson, 1993c, 194.

\textsuperscript{811} Rawlings, 2003.

\textsuperscript{812} Brandl, 1993.
predicates are true of them. Doesn’t this imply that a mental event is a change in the conditions under which a mental predicate can be applied to a person? If this is so, it is certainly possible that the mental properties of a person can change without any change taking place in her brain. Let us suppose that I have a specific belief \( C \) which, according to Davidson, means that I have a certain property. Suppose that a mental event occurs, i.e. the specific belief \( C \) changes to another belief \( C_1 \). If the truth of O-physicalism is accepted, there would be a change in my brain state, because my conscious state would change to another conscious state. From an “ontological point of view”, the whole story would not be told by saying that the conditions under which a mental predicate can be applied to me have changed. On the contrary, the application conditions of a predicate need not change at all, and yet a mental change may have occurred in me, or more specifically, in my brain. To be exact, the total application conditions of a predicate would have changed, because the state of my brain would have changed; but this condition cannot be a relevant condition for the interpreter operating from a third-person view.

Von Wright would claim that to say that a mental change would have occurred in me or in my brain is a confusing statement. He argues: “In the brain material processes go on. There is no ‘room’ for anything mental. The subject has the sensation. Nothing mental is ‘in’ him.” With respect to sensations, von Wright concludes that sensation itself is not a state, although having a sensation is being in a state. With respect to belief, von Wright is even more wary when he wonders: “[…] is belief a ‘state’? Is it even ‘mental’? It may be said in reply: *faute de mieux* we call it a ‘mental state’. It is surely not a ‘physical phenomenon.’” I think von Wright’s discussion suffers from the same flaw as Davidson’s: both philosophers do not clearly describe what they mean by “state” or with other relevant concepts that they use. If they are following some standard philosophical usage, they do not mention it. Von Wright’s case is even more complicated than Davidson’s, because he leaves the nature of the “mental” quite unclarified. What is worthy of praise in von Wright’s approach is his reference to the philosopher’s own conceptual intuitions as well as his conviction that philosophy should not be in conflict with common sense and everyday language.

Above all else, von Wright seems to resist the desire to “substantialise” belief or other mental phenomena. He thinks that mental phenomena should not be understood as being something thing-like, but it is hard to see how this tendency could be ignored if, as von Wright’s monism supposes, mental states are states of the brain. Of course, if “…To have a

---

814 Von Wright, 1998, 118.
belief is, in some ways, more like something we do than something we… experience”\textsuperscript{815} then there is no reason to think that there has to be state of believing, a brain state of which I am aware when I have a belief. But is it correct to claim that to have a belief is not something that we experience? If I have a belief\textsubscript{c} of which I am aware of, in what sense is this not an experience of mine? To have a belief\textsubscript{c} is not anything I do, and it is not to be defined in terms of what this belief\textsubscript{c} is disposed to cause. It is simply my awareness of this belief\textsubscript{c}, which establishes for me the existence of the belief\textsubscript{c} – and this awareness is surely something that I experience. It seems to me that von Wright admits this when he notes that: “Is there… such thing as a ‘belief feeling’ a touch which singles out a certain mental phenomenon as a belief? [...] perhaps there is, in some cases, such a feeling, introspectively recognizable and susceptible of a phenomenological description.”\textsuperscript{816} I think von Wright reaches the right conclusion when he notes that:

The mental phenomena or “Cartesian thoughts, i.e. our sensations, beliefs, desires, and volitions, the reasons we have for our actions are how we, as subjects, experience that which happens in our neural system (brain)…Our own sensations, thoughts, etc. are… experiences we have….”\textsuperscript{817}

Here von Wright brings together various kinds of mental phenomena, sensations, beliefs, reasons. To define reason as an experience is a promising strategy in the attempt to solve the problem of mental causation – as I will argue in section 4.2.2. In another passage, von Wright describes the subjective experience of what happens in the brain as “a perceptualization of what goes on in the brain.”\textsuperscript{818} This perceptualization is direct or immediate. Yet, we do not observe the neural states as neural; we observe these states as mental.\textsuperscript{819} This is what grants autonomy to a mentalistic conceptualization. This sounds a more plausible understanding of the nature of mental phenomena than Davidson’s. Or let us say at the very least that it better captures our intuitions, which may of course lead us astray.

There is nevertheless a hint of constructivism or antirealism in von Wright’s views as well, following from his interpretationism. According to him, the mind–body distinction, or the distinction between the mental and the physical, is a distinction “between two ways of

\textsuperscript{815} Ibid.
\textsuperscript{816} Von Wright, 1998, 119-120.
\textsuperscript{817} Von Wright, 1994, 148.
\textsuperscript{818} Von Wright, UPc, 4.
\textsuperscript{819} In order to prevent misunderstanding, it needs to be emphasized that von Wright’s views are not always clear with respect to the question of whether we observe the states of brain when we have subjective experiences. However, it is plausible to claim that we do not perceptualize these states as neural, because in order to do that we would have to literally see our brains. We could see these states as neural, for example, by following a brain-surgery of our brain but this obviously is something different than our normal way of experiencing these states.
looking at living beings.”

This resembles Davidson’s claim that the mental is a conceptual category, one way to see and describe the behavior of fellow humans. According to von Wright, attributing intentions to a creature is a way to conceptualize its reactions in a certain way. As he says: “When movements in or of the body of a living being are understood or ‘seen’ or described as intentional, I shall say that they are being conceptualized under the aspect of intentionality.” Sometimes von Wright refers to this aspect with an interesting term “aspect of spirituality”. Bodily movements, because they are conceptualized in a certain way, are understood as an action of an agent. Von Wright notes: “When understanding the movements to ‘mean’ an intentional action we, as it were, ‘ascend’ or ‘leap’ from the world of matter to the world of mind (the ‘spirit’).” The realm of the mental is constituted through certain kinds of conceptualizations. What happens is a “physical thing”, a movement, but we have made a transition to the “world of the spirit”, since we have understood the physical thing in a certain way. I think that this leap to the world of spirit can also be understood by considering the “dualistic” interpretation of von Wright’s emergentism. I showed that von Wright likens emergence to a transition from quantity to quality. Perhaps the transition from matter to the world of spirit could be thought of as this kind of mysterious leap.

Von Wright shares with Davidson the view that any attribution of mental phenomena cannot be separated from the contexts in which these attributions are made. What also sounds very similar to Davidson’s holistic view is von Wright’s claim that: “In attributing reasons for action to an agent we normally also attribute to him various abilities, beliefs, desires and inclinations, the understanding of institutions and practices of the community…” as well as his view that “We attribute to a person a belief on the basis of a complex pattern of his bodily reactions (and ‘dispositions’) to react. This is a conceptualization of observations on physical phenomena under the aspect of intentionality.” The first quote fits well with Davidson’s view about the properties of mental states, and the second quote agrees with his view of radical interpretation. Davidson and von Wright agree that mental ascriptions are strongly context-dependent and therefore always open to reinterpretation, and in this sense to “non-factuality.”

---

821 As I have argued, this is precisely how some critics interpret Davidson’s position and therefore describe it as a form of non-descriptivism or antirealism. As Klagge (1990, 344, my emphasis) claims: “Davidson seems to hold that mental judgments embody a way of seeing certain beings, rather than reflecting something in them…that non-descriptivism constitutes a form of anti-realism seems fairly uncontroversial.”
823 Von Wright, 1998, 106.
824 Von Wright, 1998, 27.
I think we can conclude that if the lesson of interpretationism is taken seriously, it follows that mental states are attributed to a person whereas, if monism is taken seriously, the physical counterparts of mental states are states of the brain. Von Wright’s description of the mental and the physical as the subjective and objective aspects of the same reality is, in my opinion, the most coherent formulation of a position which combines interpretationism and monism. Interpretationism is an acceptable description of the way that mental states are attributed to others, but monism is what matters if we think of our own thoughts as ways in which we experience what goes on in our brains. I claimed in section 2.4.1 (in the discussion about the nature of token-identities) and in section 2.4.3 (in the discussion about token-reductions) that Davidson goes too far in his conclusions concerning what interpretationism implies about the nature of mental phenomena. He writes, for example, that “no interesting question about the location of the event remains that is not answered by knowing where the person was when the event occurred” or “we have no reason to locate mental states more precisely than by identifying a person for, more than this would normally be irrelevant for individuation.” Von Wright makes a similar point: “The question ‘Where are the sensations?’ can only mean ‘Where are you at the time when you have them?’ And you are where your body is, i.e. at a certain place in space.” I believe we have no reason to think that these claims are correct. Where in the brain a belief is can certainly be an interesting question. Some think that this question is not interesting at all because if mental phenomena occur in space, they most likely occur somewhere “north of the neck” as Fodor so eloquently puts it. But why should one think that the exact location of thoughts is an uninteresting question? For a physicalist, it could be a very relevant question, even more so for a reductive physicalist. If it is an interesting question we do have reasons to try to locate mental states beyond the person who is experiencing them. Why would the individuation beyond the person be irrelevant? A bold claim is made when Davidson says that, when there is a change in belief:

The relevant entity that changes is the person, and there seems no difficulty in supposing these changes

---

826 Which is not to say that this formulation is without problems. Wittgenstein (1958, 8) warns that if we are considering whether there are one or two things in the case where a person sees her brain and entertains a thought, we should not say “that he is observing one thing both from the inside and from the outside; for this does not remove the difficulty.”

827 Davidson, 1969, 176.

828 Ibid.

829 Von Wright, UPg, 25.

830 See Fodor, 1999.
Here we see how Davidson claims that a mental event is a “global change”. Yet, according to Davidson, mental events are also neural – they are physical properties of the brain. Why think that when a mental event occurs, the relevant entity that changes is the person? What can this claim about relevance even mean? If, for example, we are interested in the question of the causal efficacy of belief \( C \), a relevant question could certainly be where in the brain belief \( C \) is located. Belief \( C \) is a mental event because you cannot have beliefs without believing \( C \) them, and believing \( C \) (unlike believing) is a process which takes time and occurs in a certain place. Davidson himself ties the causal efficacy of beliefs to the fact that they are states of the physical body. He also claims that with respect to the question about the location of an event, “the location of the event at the moment is the location of the smallest part of the substance a change in which is identical with the event.”\(^{832}\) By the supposition that O-physicalism is true and mental-states are identical with brain states (which are literally composed of physical substance), it would seem to follow that a mental event is identical with the smallest part of the physical substance (the brain) in which the event (change) takes place. If we accept Davidson’s suggestion that a mental event is a global change that applies to a whole person, what then is the smallest part of the substance, or what in fact is the “substance” to which we are referring? Is it the person? What could convince us that there is no smaller part of the physical substance in which a change is an event than the person whose change it is? Indeed, as Davidson himself notes, it is an error to think that “if an event is a change in a substance, the location of the event is the entire space occupied by the substance.”\(^{833}\) So, it would seem to follow that if an event is a change in a person, then it is an error to think that the location of the event is the entire space occupied by the person.

Insofar as beliefs influence behavior it is reasonable to think, again supposing that the truth of O-physicalism is accepted, that they are states of the brain, not states of the whole physical body of a human being. It simply is unclear how the global character and the interpretation-dependence of mental states go together with Davidson’s view that: “To have or acquire a propositional attitude… is to have or acquire a property in as objective sense as can be, like a battery having or acquiring a charge or a car a coat of paint.”\(^{834}\) Of course it may turn out that beliefs cannot ever be found inside the brain as concrete entities that can be

---

831 Davidson, 1999, 655.
832 Davidson, 1969, 176.
833 Ibid.
834 Davidson, 1999a, 595.
clearly individuated. But if this turns out to be the case, or so I would contend, it is certainly not because “[…] mental events… are global changes that apply to the whole person” as Davidson argues.\textsuperscript{835} His conviction that this is how we should talk about mental phenomena is not an argument against the possibility of token reducibility. As I argued in section 2.4.3, Davidson and Rorty do not provide any real argument for their claim that beliefs are not little entities lodged inside the brain. Davidson merely notes that thoughts are not independent atoms. But this is meant as a comment about the mental nature of thoughts; I believe that, as far as neurophysical details are concerned, thoughts may turn out to be the brain’s “independent atoms”. Perhaps one could have thought that there might be principled obstacles preventing actual token-reductions, obstacles which follow from the mental nature of thoughts. But, apparently, this is not so, because it is not the case that: “there is some special difficulty about identifying the appropriate location for such [mental] events.”\textsuperscript{836} Perhaps one could have thought that the reason is precisely the difficulty in individuation. Perhaps holism, as Rorty seems to think, could have been offered as an example of a difficulty. But these difficulties are not the reason why individuation is problematic; the real reason for positing the global nature of mental events is the fact that “the only thing there is that changes when our attitudes change is us.”\textsuperscript{837} I think there is no reason to take this claim seriously. There is no reason to think that the only thing that changes when our attitudes change is us. On the contrary, this claim is obscure.

2.6 An alternative to physicalism, dualism and non-reductive physicalism

In this chapter I have strongly emphasized the physicalism of Davidson and von Wright. I have argued that, at least when seen from a historical perspective, AM was meant to be a contribution to a robust physicalism. It is no wonder that Davidson is often interpreted as continuing Quine’s work in the philosophy of mind. Although, as I have shown, Quine came to accept AM and the consequent irreducibility of the mental, his philosophy of mind is generally recognized as having a reductive or eliminative spirit. I believe that there is a trace of this spirit left also in Davidson’s work; he shows a bias towards materialism or physicalism by identifying the substance of reality with physical substance. I have argued that also von Wright shows a bias towards physicalism because any alternatives would go contrary to a

\textsuperscript{835} Davidson, 1999, 655.
\textsuperscript{836} Ibid.
\textsuperscript{837} Ibid. My emphasis.
scientific picture of the world. He goes even so far as to admit that his position bears a resemblance to eliminative materialism. I have wanted to emphasize the physicalism of von Wright and Davidson in order to discuss whether and how they could face the major criticism which is raised against non-reductive physicalism. Another reason to discuss the physicalism of Davidson and von Wright in such detail has been to consider whether they offer convincing reasons to believe the claim that everything is physical. My claim is that their reasons are not convincing. Their final position seems to be that mental–physical token-identities cannot be established. This kind of view about identities which cannot ever be established is not a contribution to physicalism; if mental and physical phenomena each exist in their own right, what is the point of claiming that everything is physical?

I suggest that Davidson and von Wright should have taken a further step in their philosophies of mind by rejecting even O-physicalism. Although I am not willing to argue in this work that a robustly dualistic ontology should be accepted, I do want to suggest that the vagueness of ontological physicalism should lead to its rejection, or at the very least, to a serious reconsideration of its status. Although dualism is treated with great suspicion in the modern philosophy of mind, there are philosophers who are willing to explore this alternative as the titles of recent books, *Psycho-Physical Dualism Today: An Interdisciplinary Approach* 838 and, *The Two Sides of Being: A Reassessment of Psycho-Physical Dualism* 839 show. The ontology of the mind is not a clearly settled issue. Kim has claimed that we have little knowledge about the possibilities and dangers which lurk in the cavern of dualism. On the other hand, he has also concluded that dualism is an uncharted territory and that a thoroughly physicalist view of the mind is not plausible. Some philosophers hope that quantum physics could perhaps provide a non-physicalistic solution to the mind–body problem. I believe that those exploring or defending a dualistic approach raise legitimate concerns against the position of physicalism. Leaving these speculations aside, I want to point out yet another ontological position which Davidson and von Wright at times seem to see as an alternative. This option is *neutral monism*.

There are clear reasons to think that physicalistic monism is the best way to describe a Davidsonian ontology. Davidson claims, for example that “psychological events simply *are* (in the sense of *are identical with*) physical events. If this is materialism, we are committed to it…."840 He notes also: "[…] psychological events are describable…In physical terms, that is,

840 Davidson, 1973, 248, second emphasis mine.
they are physical events… this position deserves to be called… monism, because it holds that psychological events are physical events….”

Finally: “Monistic my view is, since it holds that mental events are physical events.”

It is therefore the case that: “Anomalous monism shows an ontological bias [by] insisting that all events are physical.” Given these statements, it is perhaps surprising that in his later writings Davidson clearly shies away from materialism or physicalism. He says: “I would now hesitate to call this [his position in “Mental Events”] materialism. All that this sort of supervenience enforces is a single ontology of objects and events, a form of monism.” Usually the supervenience of the mental on the physical is understood as implying at least a weak commitment to materialism. But here, Davidson seems to think that his version of supervenience does not imply materialism in any interesting sense. The main reason why Davidson wants to distance his view from physicalism is the fact that his position is a non-reductive one. Davidson notes: “I have resisted calling my position either materialism or physicalism because… I do not think that mental properties (or predicates) are reducible to physical properties (or predicates)…” In his later comments, Davidson in fact firmly denies that he is a physicalist. This certainly distinguishes him from Quine and from many contemporary philosophers of mind. As Davidson says, adding to the reasons why he is not a physicalist: “I am certainly not now what people usually call a physicalist. Anomalous monism is not a form of physicalism or materialism, since it denies that physics exhausts what there is to say about the world.” Davidson could be interpreted as a true anti-physicalist because of his view that physics does not exhaust what there is to say about the world. On the other hand, I have shown that sometimes Davidson shows a clear bias towards the explanations of physics.

But there is also another interesting reason why Davidson could be seen as an anti-physicalist. Davidson could be interpreted as defending a view according to which the status of events, which are basic ingredients of reality, is in some important sense neutral. On many occasions Davidson compares his approach Spinoza’s position, which is usually seen as a form of neutral monism. Although some may not take seriously Davidson’s references to Spinoza, the extent in which he used Spinoza as a reference point is notable and therefore interesting. In the contemporary philosophy of mind, Spinoza is usually not the first philosopher to whom one refers when describing his views about the relation of mind and

---

841 Davidson, 1974b, 231.
842 Davidson, 1985, 245, my emphasis.
844 Davidson, 1974c, 281, fn. 1.The footnote is added in the year 2000.
845 Davidson, 1985, 245.
846 In Gluer, 1995, 76.
body. The reference to neutral monism raises the question of the extent to which Davidson is committed to O-physicalism. Neutral monism is clearly something different than physicalistic monism. The former denies that reality is “based” in either the physical or the mental; neither has priority. Reality is rather based in one particular kind of substance that can be classified as neutral. The latter holds that only the physical is real, or that it is somehow “more real” than the mental. The attribution of physicalist monism to Davidson has lead commentators to conclude that AM is a form of eliminativism since the “physical furniture of the world” is more real than the mental descriptions of this furniture.

Neutral monism is not a very popular position today, but it is a position which has been considered to be a reasonable alternative to idealism or materialism in the history of philosophy. Von Wright makes an interesting observation when he notes: “[...] early logical positivism... had to defend itself both against materialist charges of idealism – and against idealist charges of materialism. The defended position usually took the form of ‘neutral stuff monism’...” Davidson was no positivist, but it is interesting that AM has often been criticized as being both an overtly physicalistic theory and not being physicalistic enough. The problem with neutral monism has however always been the question of how the nature of “neutral” substance should be understood. This substance can be described in two different and irreducible ways. Can anything more be said of it?

When we consider this question, the ways in which Davidson refers to Spinoza are worth emphasizing. He says: “Spinoza was what is sometimes called a dual-aspect monist, and so was I.” Moreover, “The result [of anomalous monism] is ontological monism coupled with conceptual dualism: this is in many ways like Spinoza’s metaphysics.” Again, more strongly: “Our grasp of reality demands two different... modes of description, explanation and prediction. Here I am a Spinozist; a single ontology, but two modes of apprehending it.” In one sense Davidson’s approach resembles that of Spinoza’s very clearly. There is only one substance that can be described in two different ways, and these descriptions cannot be reduced in favor of or to each other. We could argue that the two modes of description stand on equal footing and this would mean either that the physical is a conceptual category in a similar sense as the mental, or that the mental is an ontological

---

847 The expression “substance” is perhaps misleading here, but I know of no better term to substitute.
848 Philosophers who consider neutral monism as a serious alternative include, among others, Russell, Mach and Kaila. More recently, the views of Nagel and Chalmers bear some relation to neutral monism.
850 Davidson, 1999h, 64.
851 Davidson, 1994a, 231.
852 Davidson, 1999n, 124. My emphasis.
category in a similar sense as the physical. According to Spinoza the substance, which can be described as physical or mental, is neutral in itself. This impression is possible to get from reading Davidson as well. Thus – as we have already seen – one commentator, Louisa Röska-Hardy writes: “What exists is non-abstract particulars…. They are neither mental nor physical in themselves….”<sup>853</sup> It seems that AM is at least compatible both with physicalism and with neutral monism. Davidson describes AM as being a version of “dual-aspect monism”. The dual-aspect theory in general, and Spinoza’s version in particular, accepts the idea that things which have physical and non-physical properties or descriptions are themselves neither physical nor non-physical. Crane, for example, claims that property dualism and double-aspect theory are reasonably seen as truly non-reductionist.<sup>854</sup> A true dual-aspect theorist would therefore be far away from physicalism, and the term non-reductive *physicalism* would perhaps be an inappropriate name for such a position.

The following kinds of statements also tend to weaken Davidson’s physicalistic position. “I see no good reason for calling all identity theories ‘materialist’; if some mental events are physical events, this makes them no more physical than mental. Identity is a symmetrical relation.”<sup>855</sup> It is not easy to understand what it means to say that, with respect to the mental and the physical, identity is a “symmetrical relation”. True, each mental event is a physical event and each physical event is obviously a physical event, but what does it mean to say that a physical event is a mental event if event’s being mental is only a matter of description? It can only mean that a physical event, which is also a mental event, is “as much” mental as it is physical. In other words, if an event can be described in both vocabularies, the question of whether the event, “in itself”, is mental or physical does not make real sense. When an event is described as mental, then it is “really” a mental event, although it is also a physical event.<sup>856</sup> When it is described as physical, then it is really a physical event. This holds for events which are both mental and physical, i.e. for events which can be given both kinds of descriptions. This view, which flirts with neutral monism, seems to suggest that we should not say that a mental event (which is also a physical event) is somehow more physical than mental. But this view is hard to understand if it is thought that what events are, or what kind of events there are, is a question of ontology, “a question of reality”, and, given Davidson’s views, a question perhaps of physics. I have shown how Davidson sometimes

---

<sup>853</sup> Röska-Hardy, 1994, 291.
<sup>854</sup> Crane, 2000.
<sup>855</sup> Davidson, 1987, 33.
<sup>856</sup> One should recall Melchert’s (1986) perspicuous description of AM.
likens “ontology” to what is described by natural science; an extreme view of this sort is Quine’s position (whose ontology continues to consist of “quarks and the like”).

If ontology is likened with what natural science says, then it should be said that events are physical and this status is stronger than the status of a mere mental description (which applies only to certain events). Every event and object is physical, whereas only some events are mental. What makes everything a candidate for a physical description if not the fact that they are physical – which means, perhaps, that they are composed in certain way from particles. They can be described as physical because they are so composed. One could argue that the composition or structure of everything is to be described by the science that studies composition and structure questions, and therefore this science, physics, has the last word on ontological questions. As Skillen has noted, according to Davidson, “things in themselves are physical.” But part of the fascination of AM is that it is open to various different interpretations. Siitonen, for example, claims that Davidson’s monism is “neither a neutral one, nor a reductionistic one – as e.g. physicalism and phenomenalism are.” His conclusion is that by representing monism, Davidson wants to avoid Quinean dualism. This is a very interesting interpretation, and I believe it agrees with my suggestion that Davidson is challenging Quine’s double standard. The interpretation made by Siitonen leaves nevertheless open the nature of Davidson’s monism; not neutral, not reductive, but what?

Perhaps it is a mistake to call all identity theories “materialist” even if they ontologically reduce the mental to the physical and even if it is the physical which determines the mental so that mental depends on the physical, but this is the way the talk about identity theories in the philosophy of mind is usually understood. After all, the issue here is ontological. Davidson himself writes: “In ontic language, mental events are identical with physical events….” I think a legitimate question is: what does “ontic language” mean for a philosopher who wholeheartedly accepts conceptual dualism? What is Davidson’s ontic language, is it the language of science? It is not entirely clear to what Davidson’s terms “ontological” or “metaphysical” refer. If we are committed to the view that all events are physical or material, then I think that there is no reason to deny that our ontology is that of a physicalist. It is evident that Davidson’s identity theory differs from the usual forms of contemporary identity theories in an interesting and important sense, as the reference to Spinoza shows. The main reason why Davidson cannot endorse a standard version of identity

857 Skillen, 1984, 523.
859 Davidson, 1995c, 207.
theory is that a typical version of an identity theory would jeopardize the idea of the anomalism of the mental. This idea is an independent view of Davidson’s, which he takes to be true and, since it is assumed to be true \textit{a priori}, a reductive version of physicalism \textit{must} be false. However, since also in Davidson’s case the mental is assumed – again \textit{a priori} – to be identical with the physical, and the physical is assumed to be the only ontological category there is, I see no real reason to label this kind of theory as non-physicalist. \textit{If formulated in the Davidsonian way}, Davidson’s position is physicalistic. It is therefore interesting that Davidson himself refers to his position as “a mild form of materialism”\textsuperscript{860} or as a version of “bland monism”.\textsuperscript{861} The standard position in philosophy has been that either you are a materialist, or you aren’t. If you think that everything is physical, then you are a materialist physicalist. If you think that at least something is not material then you are something other than a materialist.

Given that physicalism is usually understood as being an “all or nothing” position, the bland middle positions look puzzling. This puzzlement is evident among the critics who have evaluated the nature and importance of AM. The confusion with respect to Davidson’s position increases thanks to views like the following: “Such a bland monism… is not apt to inspire that nothing-but reflex (‘Conceiving the \textit{Art of Fugue} was nothing but a complex neural event’….)”\textsuperscript{862} Does this mean that my conceiving is something other than the neural event which is identical with my conceiving, that is, which \textit{is} my conceiving? I think that Davidson wants to emphasize that my conceiving is not “nothing but” – in the sense that we cannot reduce my conceiving of the \textit{Art of Fugue} to the neural event which is identical with my conceiving it, and that we cannot explain everything (perhaps hardly anything) about my \textit{conceiving} understood as a mental event if we understand it as a complex neural event (which it nevertheless is). But to draw strong ontological consequences from this insight would be strange, because it is already admitted that conceiving the \textit{Art of Fugue} is a complex neural event. In a sense, an ontological conclusion is what Davidson nevertheless draws. In later writings, he does not want to see AM even as a mild form of materialism:

\begin{quote}
[my position] is a form of monism. The objects and events are sometimes described in physical terms, sometimes in physiological terms, sometimes in psychological terms: the same \textit{things} are described now as physical, again as mental or biological, etc. The only reason to call this physicalism would be if one accepts the additional claim that ideal physics is more precise or complete than any other system of
\end{quote}

\textsuperscript{860} Davidson, 1974c, 281.

\textsuperscript{861} Davidson, 1970, 214. David Lewis (1974) seems to agree. He describes Davidson’s position as a form of “minimal materialism.” This view is shared among many subsequent commentators.

\textsuperscript{862} Davidson, 1970, 214.
description and / or one thinks definitional or nomological reduction of the vocabularies of psychology to the vocabulary of physics is possible. But why believe this? Here we again get the impression that the descriptions in different vocabularies stand on the same level. It sounds like events have physical or biological or mental descriptions and none is given priority in answering the question “What events really are?” Does this mean that the physical loses its status as an ontological category? Davidson is here speaking of a single ontology of objects and events, which he describes as a form of monism. He does not describe it as a form of physicalistic monism. This being the case, Davidson seems to have changed his mind. Earlier he described the position as a form of monism because it holds that mental events are physical events. His last view on the subject emphasized a single ontology of events and objects, thus the monism. This is interesting because it could really point to the conclusion that Davidson takes events to be something neutral, and that their status as “physical” is only a matter of description. We already saw how it is indeed the case that events are mental or physical only as described. One could say that the nature of reality is such that it is amenable to these very different descriptions. But there is still a sense in which mental events are more conceptual than physical events. Davidson namely argues that any event can be given a mental description, and in that sense every event is mental as described, but it is physical ontologically speaking. We are told:

Take some event one would intuitively accept as physical, let’s say the collision of two stars in distant space. There must be a physical predicate ‘Px’ true of this collision, and of others, but true of only this at the time it occurred. This particular time… may be pinpointed as the same time that Jones notices that a pencil starts to roll across his desk. The distant stellar collision is thus the event x such that Px and x is simultaneous with Jones’s noticing that a pencil starts to roll across his desk. The collision has now been picked out by a mental description and must be counted as a mental event.

We could describe any event mental in this way, but in that case the event, say the distant star collision mentioned, would surely be mental only as described. As Davidson correctly observes, this “Spinozistic extravagance” does not capture the “intuitive” concept of the mental. What this example shows, however, is a strong sense for the expression “events are mental only as described”. It sounds plausible that a collision of two distant stars counts as a mental event only because we have given it a description according to which it is simultaneous with the noticing of Jones. This property of the collision does not exist.

---

863 Davidson, 1974c, 281-282, fn.1.
864 Davidson, 1970, 211.
independently of us, whereas some other properties of the collision could be taken to be there independently of our descriptions.

The exact status of the mental and physical is still somewhat open. On the one hand we have very good reasons, based on textual evidence, to think that Davidson’s version of monism is physicalistic. However, what complicates things and increases the plausibility of the neutral monism option are remarks Davidson’s makes like: “Mental events are… physical (which is not, of course, to say that they are not mental).” An interpretation by Paolo Leonardi emphasizes that an event’s being mental or physical is something that depends only on the vocabulary in which it is described. He quotes Davidson as saying: “[…] with respect to those physical entities that are also mental entities, one could as well say that they are nothing but mental.” Leonardi claims that Davidson seems to have the view, which we have already discussed, that events are mental or physical, only as described. This being the case, ontology is relative to the theory by means of which we choose to analyze it. Our bias of attributing a special reality to a physical description, positing physical entities and privileging physical properties, depends just on the fact that physics is the core theory of what there is. But if it is the case that physics tells us what there is and the question of ontology is just the very question of “what there is”, then why shouldn’t we say that physics and thereby the “physical” are ontologically privileged?

I don’t think that an interpretation according to which Davidson is arguing in favor of neutral monism can be correct in the end. Here we must be careful: the interpretation cannot be correct based on what Davidson has said. This obviously does not mean that neutral monism could not be a correct description of the metaphysical nature of reality. Perhaps something like neutral monism is a reasonable option over physicalism, and perhaps Davidson should have emphasized this. In fact in the most promising, although somewhat superficial, study of the similarities between Davidson’s and Spinoza’s views, van der Burg notes the following:

865 Davidson, 1994a, 231.
866 Leonardi, 1999, 117, my emphasis. Leonardi quotes Davidson’s unpublished manuscript ”The Measure of the Mental”, which I haven’t been able to see. It is quite surprising that Davidson really says what Leonardi quotes him as saying. It is difficult to evaluate the meaning of the expression out of the context in which it is given. However, if Davidson really has the opinion that we could say that events are nothing but mental (which sounds like idealism or panpsychism), it is not clear how his position should be understood. Of course, this comment can also be given a “harmless reading” in the spirit of Melchert’s (1986) suggestion that a mental event is really a mental event, and it makes no sense to ask further questions outside descriptions.
867 As I showed in section 2.1.2, he is not the only commentator who makes this interpretation.
868 It is surprising that so many commentators stress the neutrality of events given the textual evidence of Davidson emphasizing that everything is physical.
Davidson identifies the physical...with substance, thus taking the symmetry suggested by the conceptual dualism out of his metaphysics. And as long as we will keep saying that all talk of the mental is merely that, a different way of talking about the more fundamental and 'real' physical world, the urge to reduce this different...talk to the realm of that more fundamental physical world is never going to be silenced. If, however, we would not identify the physical with substance this reductionist urge would no longer be so natural. 

In essential respects I completely agree with van der Burg’s interpretations. He provides a perceptive analysis of Davidson and Spinoza, and suggests that a crucial problem in Davidson’s “Spinozism” is precisely that whereas for Spinoza substance is neutral in its relation to the attributes which are used to describe it, it is equated with the physical in Davidson’s case. Van der Burg claims:

[...] it is clear that Davidson’s version of substance monism and conceptual dualism is a materialistic view... there is only one substance and that substance is matter. But this is not the case for Spinoza. He is a substance monist and a conceptual pluralist, but his monism is not materialistic.

It seems to me that van der Burg would agree with my interpretation, according to which those commentators who emphasize the “neutral nature” of Davidson’s ontology are mistaken. On the other hand, van der Burg does not refer to Davidson’s views, which I have described, in which he explicitly suggests that his monism is not materialistic. I assume that van der Burg would agree with my claim that these remarks cannot be taken too seriously; in the end, Davidson’s view is materialistic. His references to Spinoza are misleading in the way suggested by van der Burg; although both Davidson and Spinoza defend a form of conceptual dualism, there are essential differences in their metaphysics.

I believe that a step toward neutral monism is something which would have been natural one for Davidson to take given his views about the nature of the mind. There is nevertheless a clear reason why this step was prevented; I agree with van der Burg also in his claim that “Davidson should not follow natural science as much as he perhaps does in his materialism.” Finally, I completely agree with van der Burg that although Davidson has the right critical attitude towards materialism, he does not quite bear it out in his philosophy. In chapter one I criticized Rorty for the fact that he sees Davidson as a pragmatist who has displayed no interest in a physicalistic ontology. I suggested that Davidson’s rejection of this ontology is not as clear as Rorty’s and van der Burg’s interpretations can be seen. Whereas Rorty argues that according to

871 Ibid.
Davidson we can understand everything naturalistically, van der Burg claims that Davidson is simply not naturalistic enough, at least when compared to Spinoza. My interpretation falls perhaps between these two views. In chapter one, I suggested that Davidson’s and von Wright’s calls for intellectual and methodological diversity in the study of human reality resemble the views of classical naturalists. Yet, I must conclude that both von Wright and Davidson are unwilling to take the step towards true conceptual plurality and the rejection of the special status granted to natural science. I believe we should let go of the view that non-materialistic explanations of phenomena are also non-naturalistic explanations.

My conclusion is that the nature of Davidson’s metaphysics cannot conclusively be fathomed. This is related to the general problem of the nature of non-reductive physicalism: Is it best understood as a physicalistic position or as a non-reductive position? Can these positions co-exist? But, although I think that non-reductivism is a good reason to see Davidson’s position as a form of non-physicalism, his case is more complex than this. This is because it is not clear how the references to Spinoza and to dual-aspect theory might best be interpreted. My suggestion that the nature of Davidson’s metaphysics cannot be conclusively decided should cast doubt upon the common interpretations, which claim that Davidson is a physicalist or eliminativist. But the claims that Davidson is a dualist or neutral monist should be treated as carefully. This is an example of a question about Davidson’s philosophy that cannot be conclusively settled. Some critics may conclude that Davidson is a physicalist while others may conclude that he is a neutral monist, and both interpretations make sense. There is no point in trying to show which option is ultimately the correct one. Given the textual evidence that I have presented, this conclusion is easily reached. It must also be accepted that perhaps Davidson just changed his mind about whether his theory was a form of physicalism or not. This being the case, all we can do is to consider the different aspects of his view and reach a conclusion; this, in essential respects, is what I have done in this chapter. The resulting view may not be Davidson’s view but my interpretation of it.

If, despite the above conclusion, we nevertheless stress the neutral monism aspect of Davidson’s position, then it bears an interesting resemblance to von Wright’s view. This fact has been neglected in the discussions of the positions of these philosophers. I think that von Wright’s description of the relation between mind and matter can clarify the question of how Davidson’s neutral monism or “dual-aspect monism”, as he himself describes it, could be understood. As I have argued, both Davidson and von Wright see the mental as a conceptual

---

category and the distinction between mind and matter is conceptual and not ontological. Incidentally, Wittgenstein once wrote: “‘Mental’ for me is not a metaphysical, but a logical, epithet.”\textsuperscript{874} I am not suggesting that von Wright and Davidson adopt their view from Wittgenstein, but the similarity between the views of these three philosophers is interesting. When emphasizing that the mental is a conceptual category, are we saying that the physical is not a conceptual category? I think that in the case of Davidson he must, at some point at least, have had a conceptual/ontological distinction in mind. If there is no such thing as an ontological category, why stress the fact that the mental is not such a category? I think that a reasonable way to interpret Davidson is to conclude that for him the physical is something “more substantial” than the mental. This is, of course, to stress the correctness of the physicalistic interpretation. The question about the nature of Davidson’s monism is interesting, not just because we would like to find a correct answer to this question but also because the nature of his monistic position has obvious consequences for important questions in the philosophy of mind. Moreover, if neutral monism turns out to be a coherent option in the philosophy of mind, then Davidson’s and von Wright’s arguments in favor of it are a good starting point for further investigations about the nature and prospects of this position. One could say that von Wright and Davidson, insofar as they argue in favor of neutral monism, are introducing a new way of seeing contemporary questions in the philosophy of mind by reviving an old position. Davidson himself argued that Spinoza can be seen as an anomalous monist; perhaps it could thus be argued that Davidson is revoking or bringing back certain intuitions of Spinoza, which should be further discussed – not merely as part of the history of philosophy, but also with having a possible relevance for modern philosophy of mind.

When describing his position von Wright writes: “I differ from behaviourism and materialism in that I cannot accept an identification of the mental with states of affairs in the material world.”\textsuperscript{875} This, I think, is true of Davidson also insofar as we think that the token–token identity theory collapses to triviality. Both philosophers share von Wright’s sympathy “with [the] monistic view of what the world is made of.”\textsuperscript{876} But what is the nature of von Wright’s monism? Is reality neither mind nor matter but something “neutral”, out of which mind and matter, the mental and the physical, are conceptual constructions? Already in his youth von Wright had thought about “monistic philosophy”, which was inspired by such philosophers as Ernst Mach and Richard Avenarius. Eino Kaila, von Wright’s teacher, was

\textsuperscript{874} Wittgenstein, 1992, 63e.  
\textsuperscript{875} Von Wright, 1998, 148.  
\textsuperscript{876} Ibid.
tempted by a form of neutral monism, and this approach inspired also his student. Kaila could never explicate his position in a completely satisfying way, and I think that the same thing can be said about von Wright.

Von Wright warns that the position of neutral monism is apt to invite criticism if we try to “concretize” the neutral substance of reality. He writes: “The ‘neutral stuff’ is not ‘a third thing’, neither material nor mental. But the view of mind and matter as two aspects of what is real is perhaps the best way to view the double relationship of mind to matter and matter to mind. This view... not entirely unlike that of Spinoza’s.”

It is remarkable that also von Wright refers to Spinoza when clarifying his own position. This is a peculiar similarity between von Wright and Davidson and a curious detail, one which sets their views apart from most modern theories of mind. I think that it is not entirely clear how the term “aspects” should be understood here, but it seems that this view of von Wright, even in its obscurity, provides also one possible way to see how Davidson’s version of monism could be understood. The mental and the physical are *conceptually intertwined* and, this being the case, there is no sharp distinction between them. Then the distinction which we desperately try to establish is, more or less, artificial. There are no priorities between the mental and the physical conceptualizations; neither is more basic and the idea of an ontological category does not make clear sense. *Mind and matter in their relations are the neutral stuff* and to think of them as categories (one ontological, one conceptual) which could be clearly separated and considered in isolation is already a mistake.

Von Wright thought that we are moving in an extremely obscure terrain when we try to clarify the mind–body relation. I have shown how he thinks that a fatal error in the modern philosophy of mind is the separation of mind and body, of the mental and physical, in terms of the mind and brain. The mind–body relationship can be seen clearly only once the Cartesian idea that mind and matter could be conceived independently of each other is rejected. As von Wright claims:

Reflecting on the ‘nature’ of the mental, on what we ‘mean’ by attributing mental states to a being, will show that the criteria of the mental are material things and processes. Conversely – reflecting on the nature of matter, of what it means to talk of properties and relations of material things, will show that the criteria of truth about matter are mental (e.g. sensations). Such reflective processes – not empirical investigations – may lead to the view that the two are conceptually ‘intertwined’ and not separable. This last would amount to a kind of ‘non-cartesian’dualism.

---

878 That the neutral stuff is mind and matter in their relations is the insight of Emiliani, 2001.
879 Von Wright, UPf, 1.
This is a very interesting view. Von Wright suggests that monism is a conceptual or metaphysical view because the monistic conclusion can be reached through "reflection". Dualism, which von Wright likens to interactionism, is an empirical position because it holds that the relation between mind and body is causal and not conceptual. In my opinion this is an intriguing and very tempting take on the mind–body problem.

The conceptual relation of mind and body is what really opens the way to neutral monism. Von Wright was troubled by the question of how the nature of this position could be clarified. One way to try to describe the position is to claim that mental and the physical are two aspects, the subjective and objective aspect, of the same. This would resemble the "dual-aspect" view. But, as von Wright asks, what is the "same" of which mental and physical are two aspects: "Shall we say reality? It is difficult to find the right word here." When contemplating the conceptual nature of neutral monism, von Wright considers whether its main message should be just that mind and matter are "conceptually intertwined" and not "conceptually separable"? He goes on to ask: "But if so, why call this monism rather than dualism"? Von Wright’s position oscillates between different alternatives, and he despairs of finding a satisfying final characterization. We hear: "Is my position to be labeled, philosophically speaking, ‘dualism’ or ‘identity-theory’? One can accept both labels – and at the same time reject both." Or: "Am I a ‘dualist’… who thinks that there are physical and mental phenomena and that the one kind is ‘irreducible’ to the other? Perhaps – but the meaning of the question is not quite clear to me." His final formulation was that the position deserves to be called "non-reductive dualism", which has to compete against reductive materialism and reductive phenomenalism.

In sections 2.2 and 2.2.1 I rejected the suggestion that von Wright subscribes to a neutral stuff monism and claimed instead that his position is actually a rather robust form of physicalism. I think there are good reasons to stand with this interpretation. But the "neutral stuff" interpretation cannot be completely rejected, and I believe that it is in fact philosophically much more interesting than von Wright’s physicalism. On the one hand, von Wright’s position definitely has important connections to physicalism, but on the other hand, in the end he came to describe his position as a form of dualism. Emiliani has tentatively labeled von Wright’s position with the title “Bi-Polar Monism." Incidentally, in 1912

881 Von Wright, UPf, 2.
882 Von Wright, 1998, 166
883 Von Wright, UPg, 32.
884 See, Emiliani, UP.
William James used the term “bilateral monism” when referring to the position of “scientific or Spinozistic positivism.”

According to Emiliani, mind and matter are opposite conceptual poles that presuppose and require one another; thus the label “Bi-Polar” for the position. Emiliani’s interpretation does justice to von Wright’s statement that “mind can be said to depend, conceptually, on matter, and matter on mind.” This observation is what justifies the position of neutral monism. But how to understand the claim that mind and matter depend conceptually on each other? I have shown that according to von Wright, mental phenomena have behavioral criteria. These criteria cannot be conclusive. Therefore in any third-person statement about the mental there is a residue of meaning which cannot be captured by behavioral criteria. This excludes the possibility of identifying the presence of mental phenomenon M with the bodily movements B. An identification of M with neural state N would be secondary to the identification of M with B. We must first know, relying on behavioral criteria, that a subject is experiencing M; it is only after this that further identifications are possible. But the identification of M with B cannot be conclusive, as von Wright notes: “Logically speaking there is always discrepancy between B and M.” The ignorance of this discrepancy is the materialist error. An attempt to identify M with P is to ignore that “Mental phenomena are in a characteristic sense transcendent in relation to material bodies and states.” The term transcendent is interesting. I believe it can be related to von Wright’s views about the nature of emergence and to his view of how quantity turns into quality.

But if M is not identical with a physical phenomenon, what is then the status of M? As von Wright asks: “What is the sensation ‘itself’ then – this mental or psychic thing? What is its ‘mode of existence’, what its ontological status?” He notes: ‘The ‘residue of meaning’… is the locus of the ‘purely mental.” Insofar as we want to talk about the location of a purely mental phenomenon, the location is “in” the residue of meaning. This again is a very intriguing suggestion, which in my view cannot be clarified much further. The transcendent nature of mental phenomena is due to the fact that they are located in the residue of meaning and that they are inaccessible to exact measurement in time. I believe it is these two properties of mental phenomena which ultimately explain the sense in which they are

885 James, 2003, 119.
886 Von Wright, 1998, 110.
887 I return to this point in section 3.2.
888 Von Wright, UPc, 4.
889 Von Wright, UPe, 1.
890 Von Wright, UPg, 29
891 Von Wright, UPh.
emergent phenomena according to von Wright. What about the *subjective aspect* of mental phenomena? The residue of meaning occurs when mental phenomena are seen from the third-person perspective. But I am aware of the mental phenomena which I experience or “have” directly or immediately; behavior plays no part in my experience. How to think about this subjectivity? Perhaps because of his Wittgensteinian sympathies, von Wright was careful not to emphasize too much the subjective aspect of mental phenomena. Such an emphasis would have brought us dangerously close to the web of mystification which philosophy has created around the inner-outer distinction of Cartesian dualism. But von Wright claims also that there is a genuine place for the subjective aspect of the mental, because of the irreducibility of subjectivity which cannot be captured from the outside. I think that the claim about the residue of meaning could be seen as an attempt to capture this very aspect. Von Wright did not use the favorite term of many modern philosophers of mind – qualia – but his view about the irreducibly subjective element can perhaps be seen as bearing some resemblance to them. The subjective side of the mental is summarized by von Wright when he notes that mental states are the *subject’s* way of experiencing his brain states.

Now that we have found a “place” for the mental, how can we explain the claim that mind depends, conceptually, on matter, and matter on mind? I am not sure whether I can do full justice to von Wright’s subtle position, but I attempt to clarify one way in which to think about the position. Von Wright’s view is that when we attribute mental phenomena to a subject, we conceptualize physical phenomena under the aspect of intentionality. But when we attribute qualities to physical phenomena, we actually conceptualize mental phenomena under the aspect of materiality. As von Wright writes: “Just as bodily movements are ‘signs’ of mental phenomena, sensations are ‘signs’ of physical phenomena. Mental things and events have behavioural criteria and material things and events sensational criteria.”

The concept of a purely mental phenomenon, one which would be completely detached from the material world, cannot be grasped. Perhaps we can imagine a “disembodied spirit” whose existence would be nothing but contemplation. Such a being would, however, be cut off from us. If we consider how we could come to know about its existence, we would have to refer to behavioral manifestations – and this in turn of course requires that the being should have a physical body. Von Wright concludes that what we mean when we say that a being was experiencing such and such mental state is that the being would have reacted in a certain way had we put it in an appropriate test. If we forget the disembodied spirit and consider the

---

human case, the “purely mental” can be captured only through a counterfactual move. Davidson claimed that: “Disembodied minds wouldn’t be interpretable by any methods I can think of” and von Wright noted that he failed to see how the criteria for mental phenomena could be anything but corporeal. These views, which emphasize the primacy of interpretation, attempt to discredit the idea of a “Cartesian lonely soul” which would be causally cut off from everything. Kim has considered the unfortunate fate of this kind of soul by drawing on the principles of physicalism and concluded that a disembodied being could not interact with the physical world. I believe the same conclusion can be reached merely by contemplating on the methods which are necessary for the attribution of mental phenomena. A disembodied being could not have the kinds of states which we call thoughts. This is not an empirical claim.

A counterfactual move is required to make sense of the purely mental. But the same seems to be true of physical phenomena. The idea of objective physical reality cannot be grasped. To say that a material object exists independently of us means, roughly, that were there a living being with sense organs sensing the object, the being would have the appropriate sensations of size, color and so on. The idea of mind-independent reality requires a similar kind of counterfactual move as was required in the attempt to hold on to the idea of a disembodied mind. A residue of meaning describes the relation of the mental to the physical, and this is what makes the mental to be transcendent in relation to the physical. But, interestingly, a similar kind of residue exists also in the relation between the physical and the mental because, as von Wright notes: “[…] all the necessary sensations (of a material thing) may be there and yet the physical phenomenon absent”. What stands in the way of clearly understanding the nature of neutral monism is “…the double dependence of matter on mind and of mind on matter in combination with the residues of meaning of the two in relation to each other…..” The two residues of meaning in relation to each other prevent the reduction of matter to mind and the reduction of mind to matter. Because the criteria of the mental are physical and the criteria of the physical are mental, mind could be said to depend, conceptually, on matter, and matter on mind. This would mean, paradoxically, that: “reality… is neither mind nor matter and that it is both mind and matter.” According to von Wright this was not necessarily an unacceptable position.

Davidson, 1999i, 481.
Von Wright, 1998, 111.
Von Wright, 1998, 121.
Von Wright, 1998, 110.
Nagel has suggested that when we try to reason about the possible relations between things, we have to rely on our conceptual grasp of them. I believe that von Wright deepened our understanding of the mind–body relation precisely through conceptual investigations. It remains to be seen whether these investigations can be carried further so that a new truly radical and illuminating position emerges as a result.

2.7 Conclusions of chapter two

In this chapter I have discussed the nature of Davidson’s and von Wright’s physicalism and considered some of the problems of their positions. The purpose has been to clarify the content of non-reductive physicalism and to consider whether this form of physicalism is warranted.

In section 2.1, the status of AM was considered by placing the theory in a historical context. Davidson’s case is especially interesting because, as I showed, AM was meant as a contribution to physicalism. I have claimed that, when it was given, AM was a relevant argument for physicalism, because at that time the consensus on physicalism was not yet in place. AM is thus an important part of the history of the philosophy of mind, although its current relevance as an argument for physicalism is questionable. This is so because, as I argued in section 2.1.3, Davidson’s *a priori* argument for physicalism is based on a principle the status of which is highly suspect. We have not found a reason to believe in PNCC. My conclusion is that Davidson’s physicalism is thus unwarranted in two ways. If Davidson accepts the truth of physicalism as a scientific truth, then there is no need for AM – as an argument for physicalism. If the scientific evidence is enough, as many contemporary philosophers would claim it is, then there is no need for an additional argument for physicalism. If Davidson does *not* accept the truth of physicalism at face value, then he needs an argument for physicalism – and this argument we have found to be unconvincing. Davidson’s overall position can be criticized also for the fact that his philosophical argument is meant to be contribution which *adds* to the *empirical* reasons to believe that physicalism is true. In doing this Davidson is somewhat guilty of providing a theory which, or at least so some would claim, should not be the task of a philosopher. I think that in trying to come up with a view like AM, Davidson is not entirely true to his Wittgensteinian conception of philosophy that I attributed to him in the first chapter. On the other hand, perhaps Davidson’s

participation in the debate which takes the truth of O-physicalism for granted shows that he was not willing to take the Wittgensteinian step as far as I would like to think.

The case of von Wright’s *unwarranted physicalism* is even simpler. I showed in section 2.2 that he accepts physicalism because an alternative would amount to a non-scientific view about the nature of reality. Whereas this is certainly an acceptable position, it shows that, contrary to Davidson, von Wright does not offer, and does not even attempt to offer, philosophical support for physicalism. If one does not accept the truth of physicalism at face value then nothing that von Wright says can convince one of its truth either. On this issue von Wright is perhaps more true to the purpose of philosophy than Davidson. He does not try to *prove* that reality is thoroughly physical but is satisfied with the observation that this is how things seem to be. An important contribution of von Wright is his discussion about the postulational nature of this assumption, and his clarification of the obscurities and conceptual confusions of the monistic or interactionist positions. I believe these clarifications are far more illuminating than materialistic views which proceed without considering these confusions; in a word, von Wright is doing better philosophy.

I have argued that, contrary to some commentators, von Wright’s physicalism bears obvious similarities to Davidson’s position – the main difference being precisely that von Wright does not attempt to give an argument for physicalism, whereas this is one of the essential purposes of AM. Von Wright has often been interpreted as having nothing to do with naturalistic theories of the mind, and it is thought that his views remain outside the contemporary debates. My suggestion, however, is that his version of physicalism is as strong or as weak as Davidson’s, and that the strength of their naturalism is likewise on the same level. My conclusion is thus that the physicalism of Davidson and von Wright is unwarranted – unless they simply accept the reductive naturalistic view that philosophy has to be continuous with science and that it is science which *really* describes what there is. Whereas it looks obvious that the physicalism of Davidson and von Wright is meant to be warranted through the acceptance of naturalism, this strategy is unsatisfactory because the status and truth of naturalism – and thus physicalism – is still under debate. Physicalism has not been shown to be true. Neither von Wright nor Davidson offers any convincing reasons, in addition to naturalism, to believe that physicalism is true. By letting science dictate the nature of our ultimate ontology von Wright and Davidson, despite their Wittgensteinian roots, come dangerously close to the position warned against by Wittgenstein; naturalism may lead us into complete darkness.
Although Davidson and von Wright do not succeed in their defense of physicalism, it is clear that they stand opposed to dualism in their attempt to be physicalists. This is especially true of Davidson, since AM was meant as an argument for physicalism. Likewise, von Wright noted that modern dualists were taking the situation back to the “neighbourhood of Descartes”, which is a place where a modern philosopher of mind definitely should not be. In the contemporary philosophy of mind, non-reductive physicalism has been criticized for the fact that it forms an incoherent combination. The claim is that non-reductive physicalism is a form of “property dualism”, and that a position which intends to be a form of robust physicalism should not tolerate non-physical properties. However, as I claimed in section 2.3, the discussion about mental and physical properties is, as a result of an improper and unclear use of terms, often confused. It is not clear that Davidson’s position can be accused of genuine property dualism, because he does not accept the kinds of mental properties whose existence his critics take for granted. In sections 2.3.4 and 2.3.5 I clarified Davidson’s views about properties. The conclusion was that his views about the status of properties are not definite, and thus different interpretations are possible. Both Davidson and von Wright nevertheless reject the possibility of type–type correlations between mental and physical phenomena.

I also noted that a non-reductive physicalist does not have to reject the view that mental properties are physical properties if the term “property” is understood in the correct way. Mental properties do not therefore pose an insuperable problem for a non-reductive physicalist, and both Davidson and von Wright could have answered the charge of property dualism without jeopardizing their views about the irreducibility of the mental. It is only the reductive physicalist view about the nature of mental properties which leads to the idea that non-reductive physicalism is committed to a dualism of properties. Von Wright and Davidson could have challenged this view about properties more strongly, for example, by drawing on Wittgenstein’s insights concerning how terms are used, how certain language games are played. We can agree with the common criticism that non-reductive physicalism is not physicalism enough – but this is not because its acceptance of non-physical properties. It is because non-reductive physicalism does not give good reasons to believe that O-physicalism is true. But the same criticism applies also to reductive physicalism.

Another problem, one which is often seen as challenge for a non-reductive physicalist, is the problem about the nature of token-identities. It was argued in section 2.1 that Davidson’s view is a form of O-physicalism, and it thus rejects non-physical substances. Von Wright has a similar view. Davidson’s argument for physicalism is meant to show that each
mental token is physical. In sections 2.4.1 and 2.4.2 I discussed Davidson’s and von Wright’s views about the nature of token-identity. My conclusion is that von Wright’s claim that the idea of token-identity is a serious confusion is better argued than Davidson’s view about identity, which is based on his unsuccessful argument for physicalism. I showed in section 2.4.3 that Davidson’s last word on the subject is skeptical. Because of his argument for monism, there have to be token-identities; and yet we may be never in a position to find them. I also noted that non-reductive physicalists in general are skeptical towards the view that actual token-reductions are possible. This is a strange conclusion if the possibility of token-reduction is considered to be an empirical issue and if the non-reductive physicalists do not give reasons, based on available empirical evidence, for why such reductions look implausible. This being said, my conclusion is that the skeptical attitude towards the possibility of token-reductions even further emphasizes the unwarranted nature of the physicalism of non-reductive physicalism. If mental to physical reductions are not forthcoming, we have no good reasons to believe that mental tokens are part of the physical world. At any rate, von Wright and Davidson fail to provide these kinds of reasons.

Whereas von Wright’s and Davidson’s arguments for physicalism are unconvincing, their views about the nature of mental phenomena are interesting. In section 2.5.1 I claimed that the most serious problem of their positions is their insufficient attention to the nature of consciousness. As I showed, this was the last topic in von Wright’s philosophy of mind. It is unfortunate that his views were left unfinished. In section 2.5.2, certain misunderstandings about Davidson’s radical interpretation were clarified and in sections 2.5.3 and 2.5.4 the possible consequences for interpretationism were suggested. I believe that these consequences capture the essence of the non-reductive side of Davidson’s and von Wright’s positions. In section 2.6 the possibility of neutral monism was tentatively explored as an alternative metaphysical framework. Perhaps a plausible framework, which would non-reductively incorporate mental phenomena into a naturalist framework, could be found from this direction. Von Wright’s views about the mutual dependence of mind and matter could provide an opening towards such a new position and should be further studied in the future.

My conclusion is that non-reductive physicalism is physicalism by name only. The lack of convincing arguments for physicalism, together with arguments against the possibility of reductions, leaves the physicalism of non-reductive physicalism unwarranted. This is not to say that non-reductive physicalists would in any way show sympathies for the idea of substance dualism. We have seen that Davidson and von Wright stand firmly against this idea. But whereas their views do not give support for substance dualism they do not give support
for physicalism either. In chapter two Davidson’s and von Wright’s general views about the ontological mind–body problem have been described and their views about the physical nature of mind have been found relatively unclear and unwarranted. We have thus answered first of the main questions of this thesis, namely: **How should we understand the physicalism of non-reductive physicalism, and is such physicalism warranted?** For the most part, the discussion of this chapter has been carried on under the assumption that O-physicalism is true. This has allowed us to consider, for example, the relation between mental and physical properties and to consider whether mental states are states of the brain. Although I have defended the view that mental phenomena are located in the brain this, or other physicalistic assumptions, are not something to which I am committed to. For the sake of argument, I have explored these questions from the perspective of O-physicalism which is a perspective accepted by von Wright and Davidson. Personally I may be tempted more by dualism, emergentism, pan-psychoism, epiphenomenalism or neutral monism than by physicalism.

In the next chapter I will describe in detail how Davidson’s and von Wright’s views about the nature of mental phenomena may explain why actual ontological or conceptual reductions of the mental to the physical are not forthcoming.

**Chapter three: The irreducibility of the mental**

In chapter two, different aspects of Davidson’s and von Wright’s solution to the ontological mind–body problem were considered. The conclusion was that their answer to the questions: “in what medium do mental phenomena take place?” and “how are they related to the physical world?” is quite straightforward. There is only one substance, so the answer to the first question is that materialism or physicalism is the correct ontology of the mind. The nature of the exact relation of mental phenomena to the physical world is less clear. Both Davidson and von Wright have non-reductive views about the mental, and they argue against type-type identity. Both reserve an important role for the concept or phenomenon of supervenience, although its explanatory import is unclear. We could conclude that supervenience merely states the nature of the irreducibility and does not further clarify the relation between mental and physical phenomena.

Davidson attempts to provide an argument which would show the identity of mental and physical tokens, whereas von Wright claims, with reservations, that the idea of this kind
of identity is more confusing than clarifying. Mental states are states of whole persons and not states of their brains. I believe that the details of Davidson’s and von Wright’s positions seriously undermine their common physicalistic position, and thus the claim that “everything is physical” is rendered hollow. I think it is worth asking why the need to establish the truth of physicalism is so important, given the simultaneous conviction that the mental is irreducible. Why is the ideology of physicalism so strong? Their attempt to establish the truth of physicalism fails, but this does not make the monism of Davidson and von Wright non-physicalistic in spirit. It is curious that in his later writings Davidson wanted firmly to distance his position from physicalism. This is interesting given that originally AM was meant as an argument for physicalism and a contribution to the physicalist project. On the one hand Davidson wants to defend physicalism; on the other hand he wants to claim that he is not a physicalist. Von Wright wants to defend materialism because the scientific world view demands it, but he also wants to point out the obscurities of the materialist position. One way to solve the inherent conflict which can be detected in the views of these philosophers is to emphasize the interpretation that their common ontology is robustly non-physicalistic. Acceptance of the ontology of neutral monism would at least partly solve the conflict. Unfortunately, this option was not taken very far by von Wright and Davidson, and therefore continuing this line of thought in this work would amount to speculation. I shall not do it despite the fact that the non-physicalistic spirit of von Wright and Davidson would be better served by a non-physicalistic ontology, and despite my own belief that a non-physical ontology makes better sense than the ontology of O-physicalism.

The discussion about the nature of physicalism and non-reductive physicalism showed that the positions of Davidson and von Wright are not as incoherent as one might assume given the interpretations of various critics. Antony, for example, has recently claimed that “a full defense of psychology… requires an account how psychology… could be embodied in material beings.” Similar statements are heard from other reductive naturalists. It is curious that Antony does not recognize the efforts of von Wright and Davidson to save psychology. Davidson’s famous paper “Psychology as Philosophy” defends the autonomy and special character of psychology, as does von Wright’s “In Defense of Psychology”. It cannot be denied that these two philosophers have done a lot of work to defend the autonomy and importance of psychology. This does not satisfy a critic like Antony who wants to defend psychology as a science, and perhaps as a competitor to a brain-based psychology or

---

899 Antony, 2007, 146.
900 See Davidson, 1974b, von Wright, UPa.
philosophy of mind. This defense is based on the acceptance of a certain kind of metaphysical framework and on the demands that this framework makes. Antony claims that even non-reductive physicalism is required to give a reductive explanation of psychological phenomena. This is because, according to Antony, a firm distinction between the epistemic grounds for our acceptance of mental concepts and the “ontological constitution” of the psychological realm must be made. Similar kinds of demands are made by other reductive naturalists. Fodor notes: “It is hard to see... how one can be a Realist about intentionality without also being... a Reductionist.”

Antony complains that those non-reductive materialists, like Davidson, who are not willing to provide a reductive explanation, are confusing the epistemic and ontological issues. According to her, there is a “scientific impulse” asking what it is in the world that makes psychological attributions true. Again, Fodor agrees when he claims that there is a “naturalistic consensus” claiming that “[...] something has to be said about the place of the semantic and the intentional in the natural order....”

In making these claims Antony and Fodor are committing simple mistakes against which some non-reductive physicalists warn. To insist that there must be a specific kind of answer to the question of what it is in the world that makes psychological attributions true is to ignore the possibility that perhaps in the case of the mental, the epistemological and ontological questions cannot be separated. Together with many reductive and eliminative materialists, Antony is making the mistake of which Wittgenstein, Malcolm and many others have warned. It is to assume that there is a “nature” or “essence” of mental phenomena to be found, and that the truth of psychological attributions does ultimately depend on the “physicalistic vindication” of mental phenomena. I think that a rejection of these considerations, i.e. a rejection of the discussion about whether these assumptions should be accepted, is a general problem for the contemporary philosophy of mind. What I have wanted to emphasize in chapter one is that, pace Antony for example, Davidson does not stay silent about the “ontological constitution of the psychological realm”. Pace Antony, Davidson is not willing to stop arguing once it has been shown that psychology is eliminable.”

He is not, pace Antony, claiming that “it cannot be a genuine fact that any physical event is identical with any mental event.” Davidson claims that mental events are identical with physical events and that the facts about them are as real as the facts about anything else. How is it

---

901 Fodor, 1987, 98.
902 Fodor, 1990, 12.
903 Antony, 2007, 146.
904 Antony, 1989, 183.
possible, then, to claim that, according to Davidson, the identity between the mental and the physical cannot be a genuine fact? It is possible only if one has a very peculiar understanding of what a genuine fact can or must be.

I have shown that Davidson and von Wright make various claims about the ontological constitution of the psychological realm. As the discussion about token-identities and token reductions shows, they have views about the “physical constitution” of the mental. As the discussion about interpretationism shows, both philosophers have also views about the “mental mode of being” of the mental. Davidson and von Wright say a lot, if we agree with Kim that with respect to the ontological problem the substantive question is whether or not things like belief, desire, emotion, and consciousness are reducible to neural, biological, and physicochemical properties and processes. I think we should indeed agree that this is the substantive question. The ontological aspect of the mind–body problem is simply the problem of whether a “thing” like belief is a neural process. Whatever Davidson and von Wright are saying about the irreducibility of the mental, they are not challenging the view that beliefs are neural processes. They are even defending the view that in some cases we could identify a mental phenomenon with its physical counterpart. Therefore there is no ontological issue with respect to the views of Davidson and von Wright. It is a misstatement to claim that Davidson’s arguments in favor of irreducibility are “[…] really echoes of the old anti-materialist arguments against the claim that mental states are physical states.” Von Wright and Davidson can be interpreted as supporting a robust form of physicalism despite their claims for the irreducibility of the mental. Their defense of the autonomy of the mental does not make their positions insufficiently naturalistic or insufficiently physicalistic.

In the works of Davidson and von Wright, the general ontological framework is established, although the status of mental phenomena is left somewhat unclear. What is essential to the position of Davidson and von Wright, in addition to their physicalism, is their conviction that mental phenomena are irreducible to their physical counterparts and especially that mental concepts are something that cannot be eliminated or replaced by physical concepts. The way in which mental phenomena are irreducible was discussed in chapter two. There it was shown that the irreducibility of phenomena is perhaps not as principled as one might have assumed from reading the views of critics. This being the case, the substantive

---

905 This claim must be prefaced with a caveat, since we have seen that both Davidson and von Wright, although accepting the token-identity thesis, sometimes challenge the view that a belief is a neural process. Interpretation-based philosophy of mind pulls towards antirealist consequences, whereas monism points towards realism.
906 Antony, 1989, 185. What are these “old anti-materialist arguments” and have they been refuted?
907 Although, as we have seen, van der Burg (2007), claims that Davidson should be more naturalistic.
question of Kim is not a challenge which would turn Davidson’s or von Wright’s positions into dualism. Davidson suggests that actual token identities could be established, although his last word on the subject was skeptical. By his use of the concept of supervenience, von Wright can be seen as suggesting that the reduction of mental states to the states of the brain, i.e. identifying the former with the latter, is at least not complete nonsense as, for example, Putnam and Malcolm claim.

In this chapter I will further clarify why and to what extent mental phenomena cannot be reduced to physical phenomena, but the focus will be on the irreducibility of mental concepts.\textsuperscript{908} The irreducibility of mental concepts is the essential point of the non-reductive position of Davidson and von Wright. It is not their insistence that mental properties are not physical properties. It is not their views about token-identities. It is therefore not “property dualism”, but conceptual dualism, which defines their position. Therefore the nature of this sort of conceptual dualism is a worth exploring. The question of this chapter is: What are the reasons for the irreducibility of the mental? The focus will again be on Davidson’s views, because he wrote on the subject much more extensively than von Wright. However, I think von Wright makes an extremely important contribution that supplements Davidson’s arguments. This is von Wright’s claim that the mental is \textit{epistemically prior} to the neural. What works as a guideline for this chapter is Davidson’s following statement:

\begin{quote}
What I have chiefly emphasized is the irreducibility of our mental concepts. They are irreducible in two senses. First, they cannot be defined in the vocabularies of the natural sciences, nor are there empirical laws linking them with physical phenomena in such a way as to make them disposable. Second, they are not an optional part of our conceptual resources. They are just as important and indispensable as our common-sense means of talking and thinking about phenomena in non-psychological ways.\textsuperscript{909}
\end{quote}

This is general statement capturing well the spirit of Davidsonian irreducibility. I believe that this spirit has been often misinterpreted. Antony, for example, thinks that the essence of Davidsonian irreducibility is the claim that “[…] intentional properties are ontologically distinct from… the non-intentional properties also possessed by their bearers.”\textsuperscript{910} But if one considers how Davidson himself describes the irreducibility of the mental in the previous

\textsuperscript{908} Perhaps one could claim that, given Davidson’s conviction that “events are mental only as described” and that “our sentences are the only measure of the mental”, a distinction between mental concepts and phenomena does not make clear sense. However, there is no reason to think that this formulation captures our intuitive conception of the mental. There are conscious mental states, and the features of these phenomena are different than the features of the concepts in terms which we describe them. In von Wright’s case the distinction between phenomena and concepts is made more strongly. As I have shown, he emphasizes the subjective aspect of “Cartesian thoughts”.

\textsuperscript{909} Davidson, 1999o, 599.

\textsuperscript{910} Antony, 1999, 37.
quote, it becomes clear that the basic charge of property dualism is misplaced; it certainly
does not capture the central point of Davidsonian irreducibility, since Davidson would not
accept the “ontological distinctiveness” of mental properties.

Although the general view about the nature of the irreducibility of the mental is clear, I
believe it has to be straightforwardly admitted that Davidson’s and von Wright’s views are
not knock-down arguments that prove the irreducibility of the mental. Philosophical
arguments rarely are. In chapter one, elements of the Davidsonian conception of the mental
were briefly discussed. In a radical interpretation, mental states are attributed according to
normative standards of the interpreter and the identity of the states so attributed depends on
their causal history and on their relations to other mental states. The defining features of the
mental realm, when the mental is understood in a Davidsonian way, are therefore normativity,
externalism, and holism. These features have an important role in the arguments for the
irreducibility of the mental. Principles constituting mental reality are also those which make
reality autonomous and not amenable to strict explanation and prediction. From this it follows
that if the Davidsonian conception of the mental is rejected, also the arguments for the
irreducibility of the mental must be rejected. It seems to me that it is a non-starter to argue
against Davidsonian irreducibility without arguing against the whole Davidsonian project for
the philosophy of mind, especially without arguing against the basic idea of interpretationism.

My view is that Davidson’s views about the irreducibility or anomalism of the mental
are the most interesting part of his philosophy of mind, because of their philosophical
consequences. Davidson has claimed that something like AM is needed if we want to make
sense of the idea of human freedom. This is a challenging claim. The anomalism of the mental
removes the mental, which is a defining feature of humans, from the domain of physical law
and explanation. The irreducibility of the mental has far reaching concrete consequences,
which relate to current discussions in which conclusions about the mind are drawn. I argue
that the irreducibility of the mental is something that should be defended. A view about
humans with irreducible mental states is something that needs to be defended in order to save
the dignity of individual human beings. The question about the irreducibility of the mental
connects concretely to the question of what kind of human image we want to hold. This is a
political, philosophical, scientific, sociological, and religious question. These are some of the
areas of human life for which this question is relevant. I hope that we would be more willing
to defend the view that human nature is not fixed, and that we would have courage to take a

911 Why this is the case will be clarified in the appendix.
stand on the wider issues which relate to the questions in the philosophy of mind. We can take a stand on these wider questions and should not leave philosophical claims hanging in the air without any connection to reality which, after all, is what in the end influences us all. What kind of conception of human being we want can show itself, for example, in the philosophical attempt to try to defend the autonomy of the mental.

What is the point of defending the irreducibility of the mental? Why insist that psychological facts are autonomous with respect to physical facts? What is the rationale behind my claim that the irreducibility of the mental should be defended? To this question, I can give only the following answer: philosophy should be partly a normative discipline; philosophizing should have consequences. What is the point of philosophy if it does not help with the problems of life? What is the point of philosophy if it does not improve your thinking about the questions of ordinary life? My claim is that when a philosopher takes a stance on a question like the question about the irreducibility of the mental, he is not only taking part in a merely technical discussion of professional philosophy. By taking a stance, such a philosopher is also committing to certain values. A philosophical position must start from the acceptance of certain basic principles, which cannot be further grounded. There are choices that cannot be proven to be correct but which must be made. This is a von Wrightian view which I accept, but why I accept it has to remain without clear explanation. One consequence of this view is that in the following discussion I am making one choice which cannot be further grounded. It is the very choice I make to defend the irreducibility of the mental. Given the view of man that I think we should hold, my starting point is biased, perhaps quite strongly. I cannot give an argument as to why a certain view of man would be better than another; but the choice has to be made nevertheless. Based on this choice, I claim that we should defend the autonomy of our common-sense understanding of ourselves against overtly scientific alternatives. Behind this choice there are certain personal fears and doubts about the consequences of technological progress. Perhaps this could be called my philosophical motivation for defending the irreducibility of the mental. I believe it would be dishonest to claim that philosophical views exist in isolation; there is always some motivation for a particular view. As Searle has noted: "Some computationalists invest an almost religious intensity into their faith that our deepest problems about the mind will have a computational solution."912 This is, of course, only one example, but some motivation explains also the computationalists’ intensity, and it seems to me that is plausible to think that this motivation

912 Searle, 1997, 190.
relates to broader, for example to political, issues. The question of whether mental phenomena can be naturalized is a philosophical question which is considered in the philosophy of mind. But many claim that this question can and should be seen in a broader setting. Fodor notes:

If it turns out that physicalization – naturalization – of intentional science… is impossible… then it seems to me that… we should stop doing intentional science and that counts a lot more than some philosopher being worried. That’s a matter of theoretical honesty. The position we’re in is, can we give a coherent account of this research project because, by Christ, if we can’t we should stop spending the taxpayers’ money.913

Fodor continues by noting that the question of whether intentional science is possible is related to pragmatic questions such as which philosophers or scientists should get what kind of grants and for what kind of research.914 Apparently, philosophical intensity is sometimes guided by very pragmatic considerations.

It is interesting that in a philosophical work one needs to defend a non-reductive position. Whereas our intuitions would perhaps say that our view about ourselves should have a priority and that a scientific understanding should try to challenge the commonsense view, the current situation seems to be the opposite. It is the non-reductive view that is required to justify its existence. Positions like Davidson’s and von Wright’s are treated with suspicion, because they do not fit well with scientific facts. Yet our lives are based on the common-sense understanding that we have of ourselves and of others, which we formulate in terms of mental concepts. We manage to live very well without any specific knowledge about the functioning of the brain. We succeed quite well, surprisingly well, in explaining and understanding ourselves and others without using neural concepts. Still, Davidson feels that he has to defend the view that mental concepts are important and indispensable. Given the importance of mental concepts for our lives, the defense of irreducibility of the mental or the need for such a defense would be very absurd if it were not for the unhealthy materialistic trend prevailing in the philosophy of mind. Recently Paul Churchland has claimed that:

The so-called propositional attitudes must be, at best, the occasional and ridiculously low-dimensional ‘projections’ of the mega-dimensional elements of the brain’s true representational vehicles… they are not the… relevant elements that drive our cognitive activity. More likely still, they don’t exist at all.915

It is scientific evidence which challenges folk-psychology. When this view starts to compete with the common-sense understanding on which our current life-practices are based, the

913 Fodor, 1990a, 202-203.
914 Interestingly, Patricia Churchland (1986) is worried about the same thing.
stakes are raised. If eliminative materialism were just a philosophical view about the relation between mental and the physical, it could be taken to be one proposal among many interesting, although outrageous, suggestions that have occurred in the history of philosophy. But contemporary eliminative materialism is not this kind of position. It is meant to be a scientific theory. If such a theory is accepted, it may have concrete consequences for our self-image. This image may have concrete consequences for human societies and social practices, and these consequences may turn out to be extremely harmful. To this aspect of physicalism I will return in the appendix of this work. Here I note merely that I find it important to defend the autonomy and importance of our common-sense view of ourselves given the current threats that this mode of understanding is facing.916

In the discussions of Davidson’s position, the irreducibility of the mental is usually understood in the original sense in which Davidson presented it in “Mental Events”. In that article the main thesis was that the mental realm is anomalous, it is not governed by strict laws. My claim is that the original thesis is only one part of Davidson’s completely non-reductivist account of the mental. My interpretation emphasizes three non-reductivist aspects of Davidson’s views. Although the original thesis of the anomalism of the mental is an important part of this trio, focusing too much on it, as most commentators do, may prevent one from seeing the overall non-reductivism of Davidson’s philosophy of mind. The editors of two recent collections of essays on Davidson see the arguments for the anomalism of the mental as being “notoriously difficult”917 and “intriguing, but notoriously elusive.”918 There are also more sympathetic commentators; van der Burg writes that on the whole he finds Davidson’s arguments convincing.919 I think we should sympathize with the critical views: the argument is obscure. Needless to say, in “Mental Events” Davidson provided grounds for these complaints by admitting that with respect to the argument for the irreducibility of the mental, nothing which he says deserves to be called a proof.

Whereas my purpose is to clarify these arguments as well as possible, I cannot help the feeling that the argument for the anomalism is not totally convincing. What I want to emphasize is that Davidson’s arguments between 1970 and 2000, his philosophy of mind as a whole, provide a non-reductivist view that was not in place when the argument of anomalism of the mental was first introduced. Davidson’s views developed in important ways. Therefore

---

917 Ludwig, 2003a, 19.
918 Ludwig and Zeglen, 1999, 11.
commentators who narrowly concentrate on “Mental Events” only focus on a very small part of a larger picture. Antony, for example, notes:

I should emphasize that I am not primarily...interested in explicating Davidson’s theory, which has undergone some potentially significant changes over the years. My aim is to examine the implications for psychology – whether ‘folk’ or scientific – of a view of the mental like the one set out in ‘Mental Events’. If Davidson no longer holds such a view, fine and good....

What is the interest in this? The intellectual curiosity of exploring a philosophical position could be one reason for this kind of view. But Antony is certainly not interested in Davidson-exegesis. She wants to argue that “Mental Events”, published twenty years before Antony’s criticism, suggests a view of the mental which has negative consequences for psychology. But since Davidson developed his position further, what is the point of taking “Mental Events” in isolation? Barry Loewer has recently noted that: “Davidson[s]... argument has been as influential as it is obscure to its conclusion.” He refers to the original argument. It is interesting to note how much discussion “Mental Events” has given rise to without commentators agreeing even on its most important claims. My interpretation may be controversial in the sense that it is possible that Davidson did not intend his position to be as non-reductive as I claim. If this is the case, my interpretation can be seen as suggesting what kind of conclusions Davidson should have drawn; I claim that he should have developed further the view of neutral monism. Philosophers who have a non-reductive view about the mind should explain why postulating material mind is objectionable and discuss more thoroughly what kind of undesirable consequences it may have. This is especially important now when those who have a reductive view about the mind are drawing conclusions which have consequences for the actual practices of human society. Rorty and Putnam have raised the kinds of questions that a non-reductive physicalist should ask. Von Wright was also aware of the problems that a reductive view about the mind may bring. His discussion about the scientistic attitude in philosophy, and particularly his writings which were meant for a wider audience, show that von Wright saw the problems of the mind in a wider perspective.

Most critics and commentators of Davidson’s philosophy of mind focus too narrowly on his arguments against strict psycho-physical laws, that is, on his arguments for the anomalism of the mental. Since the publication of “Mental Events”, the majority opinion has been that Davidson’s reasons are not convincing or that the view is an obscure a priori argument against the possibility of such laws. It is true that Davidson is trying to provide an a

920 Antony, 1991, 301.
argument. This argument can be accepted only if the Davidsonian view about the mental is accepted. It is easy to see that the argument does not move those who reject the kind of view of mental states that Davidson is defending. Commentators try to argue that Davidson has not offered convincing reasons to believe that there could not be any strict psychophysical laws. Whether there are convincing reasons for Davidson’s particular position or not is one question. Whether there are good reasons for supporting a general non-reductivism about the mental is another. Those who focus on criticizing the claim that there cannot be strict psychophysical laws usually ignore Davidson’s views about the irreducibility of the mental. I think that many commentators are simply not aware of the arguments for irreducibility because they have not studied the views of Davidson carefully enough. It is disappointing to find article after article where the main critical argument can be simply defused by referring to arguments that Davidson has made in papers which the critic has not studied. The issue is not, however, one of who has studied Davidson deeply enough. The main issue is that there are two different questions to be discussed. First, we must ask whether strict psycho-physical laws are possible. Second, we must determine whether the mental is irreducible, especially whether our common-sense conception of ourselves is such that it cannot be reduced to, or replaced by, a different conception.

To these two questions there are three kinds of reactions from the commentators. First, there are those who focus on the thesis of anomalism as if it exhausts Davidson’s views on the irreducibility of the mental. These reactions are a consequence of not considering Davidson’s philosophy as a whole. Second, there are those who are not interested in the irreducibility arguments at all, because they do not share Davidson’s views about the nature of mental phenomena. Third, there are those who have a generally sympathetic view towards Davidson’s approach. I believe that if we want to try to describe accurately what Davidson argues and if we want to understand the points being made, we have to evaluate his arguments in the framework of his views, and not in some preferred framework of our own. Only then can the views be genuinely criticized. Those who ignore the Davidsonian conception of the mental but keep on criticizing his arguments for the irreducibility of the mental make the same kind of mistake as those who ignore Davidson’s views on properties and keep on insisting that Davidson does not provide a reductive explanation of mental properties.

The three arguments for general non-reductivism that will be presented in this chapter lead us to see how Davidson’s philosophy is in sharp contrast to most contemporary

922 It is remarkable how many pages have been written on, for example, the questions of what Davidson means by law and how the argument against strict-psychophysical laws should be understood.
approaches in the philosophy of mind. What I emphasize is the *irreducibility* of the mental, of which the lawlessness of the mental is only one part. The exact content of this irreducibility will become clearer along with the arguments for it. Given what we already know or think we know about reality we may, as the reductive naturalists do, have the view that a complete physicalistic explanation of everything *must* be possible. It is telling that in one of the last papers Davidson ever wrote, he comments:

> There is nothing here (we say to ourselves), except physical stuff, so why shouldn’t we be able to consolidate our vocabularies for explaining, describing, and predicting what goes on? I think this is an irresistible ambition, but I am here to resist it.  

Davidson notes that at times we may feel that a reduction *must* be possible, and yet he is willing to resist the urge. This is an interesting remark if, as I suggested earlier, we want to consider the *motivation* for defending irreducibility. Davidson does not state why he wants to resist the “irresistible ambition”; in the appendix I offer my brief analysis what his motivation for this kind of resistance could be.

In this chapter I will give reasons why the non-reductive status of the mental shows that eliminativism or reductive materialism are implausible views. I claim that conceptual dualism is necessary as long as we *want* to see ourselves in a certain way, namely as rational animals who can freely act out of reasons, who are responsible for their actions and are thus subject to moral evaluation. Elimination of the mental, if pushed to extremes, would lead to a situation where we would have to change our conception of ourselves. An eliminativist could argue that we can hold whatever kind of conception of ourselves as we want, but that our *conception* of ourselves should not distort the question what kind of beings we “really” are. Since I accept this view, at least for the sake of argument, I argue that our self-conception is something which cannot be rejected, and therefore the question of whether our “real nature” can be something else than what we take it to be does not make good sense.

The trio of themes around a non-reductivist conception of the mental consists of the following:

1) Anomalism of the mental
2) Irreducibility of the mental
3) Inescapability of the mental

---

923 Davidson, 2003a, 697.
The arguments that will be presented to support my interpretation overlap each other. Together they establish the general irreducibility of the mental and suggest what the most important consequences of this view are. In this chapter, I will explicate the arguments for the anomalism and irreducibility of the mental. The case for the “inescapability of the mental”, which is not a clearly argued view of Davidson but my interpretation of how Davidson’s motivation could be understood, will be presented in the appendix.924 Von Wright’s views will be referred to whenever they provide support for the non-reductive view.

3.1 Anomalism of the mental

I will first consider Davidson’s original argument for the anomalism of the mental. The thesis of the anomalism of the mental is: there are no strict deterministic laws on the basis of which mental events can be predicted and explained.925 My discussion of the lawlessness of the mental focuses on the view that there cannot be strict psycho-physical laws in this domain. The reason for this is that the main significance of the anomalism of the mental for the questions about the irreducibility of the mental lies in the denial of the possibility of a nomological reduction between the mental and the physical. Before turning to Davidson’s reasons for the anomalism of the mental, his understanding of laws must be clarified. We have to be clear also about the question of what the thesis actually denies. The latter question has not always been clear for the commentators, and even Davidson himself has presented confusing and conflicting statements about the nature of the anomalism.

Davidson’s distinction between generalizations and laws and between strict and non-strict laws has an important role in his argument for anomalism. Laws, distinguished from generalizations, “must be true universally quantified statements. They also must be lawlike: they must support counterfactuals, and be confirmed by their instances.”926 In “Mental Events”, Davidson makes a distinction between a homonomic and a heteronomic generalization. Positive instances of homonomic generalizations give a reason to believe that the generalization could be improved by adding further provisos and conditions stated in the same general vocabulary as the original generalization. The instantiation of a heteronomic generalization may give a reason to believe that there is an underlying precise law at work, but this law can be stated only by shifting to a different vocabulary. According to Davidson...
the “laws” connecting mental and physical events have the status of heteronomic
generalizations. He admits that there may be true general statements which have the logical
form of a law, but are not lawlike. Heteronomic generalizations, or psycho-physical laws, are
at best irreducibly statistical generalizations resisting improvement without limit. Homonomic
generalizations, on the other hand, are found in the physical sciences. The generalizations of
these sciences are such that if evidence supports them, there is a reason to believe that they
may be sharpened indefinitely by drawing upon physical concepts. These generalizations
point to the form and vocabulary of a finished law.

Davidson notes that he does not object if some “important regularities” are to be
described as laws. Strictly speaking, however, regularities count as laws only if they satisfy
the conditions defined above. Heteronomic generalizations are not lawlike whereas
homonomic generalizations are. When discussing the nature of psycho-physical “laws”, this
must be remembered. Another important claim of Davidson is that regularities and laws
should be distinguished from strict laws. A strict law “should contain no singular terms
referring to particular objects, locations or times. Strictly lawlike statements do not contain
open ended phrases like ‘other things being equal’, or ‘under normal conditions’.”927 Strict
laws are generalizations which are not only “law-like” and true, but as deterministic as nature
can be found to be, and which therefore treat reality as a closed system. Strict laws do not
deploy disposition terms; nor do they use causal concepts. Davidson’s view is thus that laws
are something else than strict laws. The question is: can there be strict laws, and if the answer
is yes, then in which vocabularies can they be formulated? An answer to the first question is
important because the anomalism of the mental is the thesis that there cannot be strict psycho-
physical laws. If it turns out that there are no strict (physical) laws at all, then the whole idea
of anomalism seems to be empty.

Davidson claims that laws of physics, or “laws of nature” as he also calls them, are as
exceptionless as nature permits, and it is only physics which treats reality as a closed system.
The laws of physics therefore qualify as strict laws. It could be said that since it is an aim of
physics to find laws as complete and precise as possible, it is not surprising that strict laws are
to be found only in “developed physics”. The purpose of physics is to find exceptionless laws
and describe every physical object and event. This is a requirement that defines the aim of
physics.928 The aim of physics is closure. Underlying this view seems to be the assumption

927 Davidson, 1995c, 204.
928 It should be emphasized that this is Davidson’s formulation of the aim of physics. It is not certain that
philosopher’s view about the aim of physics is in harmony with the physicist’s view.
that reality is, in some important sense, a deterministic system. What justifies this view? I suggest that it is just a physicalistic vision which Davidson, notwithstanding his anti-
physicalist remarks, seems to have. As I noted in chapter two, Davidson believes in
“determinism” but does not really give a reason for this belief. There is no textual evidence
warranting the conclusion that Davidson’s reasons to believe in the completeness of physics
are the familiar reasons given by physicalists.

Davidson argues that whenever something causes something else, there must be a
strict law that covers the particular case. In “Mental Events” Davidson treated this principle,
which he called “The Principle of the Nomological Character of Causality” (PNNC), as an
assumption. As my discussion in chapter two showed, Davidson has later explained that
this view is a “conceptual truth”. The concepts of “law” and “cause” are conceptually related
and therefore the argument for PNCC is in some sense a priori. The acceptance of this
conceptual truth does not amount to assuming the “regularity of nature”. I think it is difficult
to say what PNCC’s exact implications for explanation and predictability are. Does a strict
law that applies to a case where A causes B explain why A caused B, and does it enable us to
predict that if A occurs, then B occurs necessarily? One natural reading of PNCC, or its
implications, is that the reality of physical events and objects is governed by strict
deterministic laws on the basis of which physical events can be predicted and explained.
Davidson claims: “Everything in the universe and its history can in principle be described in
the language of physics...” He notes that the natural sciences may arrive at a point where
they can fully explain and therefore predict any event. But on what is this bold assumption
based, and what does it actually mean? A plausible answer is that this kind of conviction
follows from Davidson’s physicalistic monism, together with his view that the PNCC is a
conceptual truth. But it seems to me unclear whether the fact that everything can be described
by physics implies that everything can be explained in the language of physics. It depends on
what is meant by the term “explanation”. Even if it were possible to describe and predict the
occurrence of any event, it is unclear whether anything would have been explained or left
unexplained. Intuitions here differ.

Although Davidson does not explicitly state what the implications of the PNCC are, I
take it to imply that if the strict law that covers the case of A causing B in a situation S were
known, the prediction that whenever an event identical to A occurs in a situation S, B will
occur too would be true. This example is artificial because there cannot be two identical

929 PNCC from now on.
930 Davidson, 1997a, 127.
occurrences of A in a situation S; the two occurrences would take place at different times. In this sense two identical events are impossible. However, since a strict law is a law which does not contain singular terms referring to particular objects, locations or times, a strict law which covers the case when A causes B in a situation S would have to state something like: If P (which is a complete physical description of A) occurs, P₁ (which is a complete physical description of B) follows. According to Davidson it is possible to describe, in principle, in the case of each physical event, why it occurred – and in this sense explain why the event occurred. Perhaps explanation in this context is not the right word, because at the most basic level we would just describe how things happen by referring to the strict laws that cover the case. In other words, let us suppose that there is a law stating that if there is a certain distribution of forces and matter in a field of a certain size at time t, it will be followed by a certain distribution of forces and matter in a field of a certain size at time t₁.³³¹ It seems to me that this kind of law does not answer the question of why a certain distribution of force and matter follows from a certain other distribution of force and matter; the law just describes that this happens and perhaps the only answer given by this kind of law is that certain things happen because this is just how things are. There is no way to answer the why-question just as there is no answer to the question why the laws of nature are what they are. What I find curious however, are Davidson’s occasional remarks that explanations get “better” when they get closer to physics. The level of strict laws is the level of perfect explanation. I agree that this would be a level of perfect prediction and description, but is this level an explanatory level at all?

The question of whether explanations in physics are better than other explanations is curious, because at times Davidson argues that all explanations are interest-relative. Or, to be more precise, sometimes he seems to argue that most explanations are interest-sensitive or relative, and sometimes he seems to claim that explanation is always geared to interest.³³² Given the latter view, the claim is that no causal explanation can be better than another independently of the context in which it is offered. So, a biological explanation could be appropriate in a certain context and an explanation in physics could be appropriate in another context; yet they could both be explaining the same event if and when the event is understood as a spatio-temporal particular. But how does this view fit with the suggestion that “if we ever had the laws of physics right, and we had the appropriate physical description of an event and

³³¹ This is Davidson’s own example of a strict law. A cruder version is that there is a strict law which says that all the objects similar to A are followed by objects similar to B. This is given in Davidson, 1967a, 160.
³³² In Davidson 1993 and 1993g, it is claimed that most explanations are interest-sensitive, in Davidson 1987 it is claimed that explanations in ultimate physics are not interest relative.
of some cross section of the preceding light cone, we might be able to give a full and sufficient explanation of the second event.”

Given that the explanation would be full and sufficient, it could exclude other kind of explanations because we would have no interest in them anymore. Davidson has a vision of “perfect explanation” which is formulated in terms strict laws of physics. This explanation is not interest relative. But doesn’t this vision reveal an unjustified bias towards the ideal explanations of physics as Davidson sees them? If one is inclined to say so, then there are two interesting additional questions. First, what is the status of interest-relative explanations when compared to the “perfect” explanations of physics? Second, what does this view tell us about Davidson’s relation to physicalism? With respect to the second question, I think that it shows how Davidson is more committed to physicalism than he himself admits. With respect to the first question, one cannot easily avoid the view that interest-relative explanations are in some sense inferior when compared to perfect explanations, which are full and sufficient.

The only strict laws are thus the laws of physics. This would seem to provide a straightforward answer to the question “Why can’t psychophysical laws be strict?” However, such a straightforward answer would make Davidson’s argument for the anomalism of the mental unnecessary. In order to solve this problem, we need to consider the argument in detail.

3.1.1 Why strict psycho-physical laws are impossible?

In order to understand why strict psycho-physical laws are impossible, it is necessary to keep in mind that for Davidson the mental is a conceptual category. The mental refers to our mental vocabulary. In chapter two I briefly described what kind of constraints Davidson takes to govern the use of this vocabulary. We attribute mental phenomena to others in a way that optimizes the rationality, truth and coherence of those attitudes. The principles of holism and normativity have an essential role in the attribution of mental states. We have seen that, according to Davidson, interpretative practices provide the only way by which beliefs can be attributed. The principles governing interpretation are constitutive principles of the mental, and in the absence of these principles the existence of the states that we have attributed to a person must be questioned. Mental reality is built on these constitutive principles in a very

---

933 Davidson, 1993d, 199.
strong way. Mental states are real physical states in a person, but we cannot make sense of the idea of their existence if they cannot be identified as mental states first.\footnote{We can ignore here the view according to which mental states are states of the whole person. The point is that mental states are attributed according to certain principles.}

The claim that mental reality is not governed by strict laws has two components. First, it is claimed that there are no strict laws that relate mental events to physical events. Second, there are no strict laws that connect mental events to each other. Together they form the general thesis of the anomalism of the mental, according to which there aren’t any kind of strict laws governing the mental realm. I will focus on the argument against strict psycho-physical laws. The main reason advanced against the possibility of strict psycho-physical laws is, as Davidson famously claims “the disparate commitments of the mental and the physical schemes.”\footnote{Davidson, 1970, 222, my emphasis.} What are these disparate commitments, commitments that cannot be compared? In Davidson’s words:

It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and conditions physically described. It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs and intentions of the individual. \textit{There cannot be tight connections between the realms if each is to retain to its proper source of evidence.}\footnote{Ibid.}

The argument in favor of the anomalism of the mental requires that the Davidsonian features of the mental and physical are accepted. The categorical difference between the mental and the physical vocabularies makes strict laws between them impossible. The main difference between these schemes is the one given above. When using our physical vocabulary we accept, or should accept, that physical changes can be described by laws which connect these changes with other physical changes. But this kind of law-involving constitutive principle \textit{is not part of our view about how our mental vocabulary is used or can be used}. It is not part of our understanding of the nature of mental reality. It should be carefully noted what Davidson is claiming: there cannot be tight connections between mental and physical realities \textit{if} both are to stay true to the evidence relevant for them.

When we use our physical vocabulary, we must operate under the constraints of that vocabulary. We are bound by these constraints if we are to use physical terms in a meaningful and systematic way. There are principles which are constitutive for our physical vocabulary, principles which govern their use. This is actually a Wittgensteinian view; it suggests that
there are constraints which cannot be violated if concepts are to be used in a meaningful way. But what is a “constitutive principle”? It could be argued that a principle has constitutive status if it in some way governs the application of concepts. A constitutive principle partly determines what counts as a correct or incorrect application of a concept. In this sense, constitutive principles play a normative role in concept application regardless of which vocabulary they belong to. Davidson does not describe in too many words the constitutive principles of our physical vocabulary, but he gives as examples the principles governing the measurement of length, mass and temperature when he writes:

[...] the whole set of axioms, laws or postulates for the measurement of length is partly constitutive of the idea of a system of macroscopic, rigid, physical objects. I suggest that the existence of lawlike statements in physical science depends upon the existence of constitutive...laws like those of measurement of length within the same conceptual domain...we cannot intelligibly assign a length to any object unless a comprehensive theory holds of objects of that sort....

In order to clarify the nature of constitutive principles, an example from “Mental Events” can be used. The claim is that we cannot consistently describe objects, for example, as having a certain temperature or rank objects in terms of their temperature unless the attributions satisfy the principle of transitivity, which is a constitutive principle of the physical. Measurements of temperature are normative because the principle of transitivity partly defines the structure of the theory that constitutes the notion of temperature. How should we react to an intransitive triad? A conclusion that the principle of transitivity is false would call into question the descriptions of an object having a temperature or being warmer than another and whether they have any clear meaning. This is what it means to say that the principle of transitivity is a constitutive principle of the concept of temperature.

Davidson argues that one important feature of physical reality is that a physical change can be explained by laws that connect it with other physical changes and conditions. This observation is vitally important. It could be taken to be a constitutive feature of physical reality, in the sense that our concepts of “physical object” and “physical event” are such that their behavior is law-governed. This is a remark on the nature of our current physical concepts; the question of whether physical concepts must obey these constitutive principles can be left open. As Davidson notes:

The causal powers of physical objects are essential to determining what sorts of objects they are by defining what sorts of changes they can undergo while remaining the same object and what sorts of changes constitute their beginnings and ends. Our concept of a physical object is the concept of an

937 Davidson, 1970, 221.
object whose changes are governed by laws... In deciding what counts as a change we also decide what generalizations to count as lawlike.\textsuperscript{938}

As I noted in chapter two, this is an interesting statement insofar as it clarifies why Davidson thinks that the PNCC is true. In “Mental Events”, the thesis was treated as an assumption, whereas Davidson later explained that the thesis is a conceptual truth and the reason for holding it must be \textit{a priori}. He writes: “If the thesis is true, what we know in advance of evidence is that if a singular causal claim is true, there is law that backs it....”\textsuperscript{939} In “Laws and Cause” Davidson returns to the reasons why he defends the PNCC. The reason is that our concept of a physical object just is a \textit{concept} of an object, the changes of which are governed by laws. Davidson claims that: “[...] to identify an object as a physical object... is already to have endowed it with certain causal dispositions; we cannot first classify an object and then discover that it has those causal properties.”\textsuperscript{940} This is what \textit{we mean} by our \textit{concept} of a \textit{physical} object; an object could not be identified as physical if its causal properties or causal behavior were totally bizarre. This is an interesting claim if we speculate on the question of what kind of discoveries may be presented by basic physics. It is unclear whether the most basic physical entities bear any resemblance to law-governed objects. As I briefly noted in chapter one, some physicists have concluded that even they cannot really understand the discoveries of quantum physics. A claim that these phenomena are part of the same physical reality assumed elsewhere in physics could therefore be taken as controversial. Perhaps Davidson’s claim that physical objects are governed by laws is appropriate at the level of macro-entities and macro-causation; the empirical investigation of the macro-level of physical reality warrants Davidson’s claim. But how far along the quantum-level can this view be applied? It is interesting that on the one hand Davidson often refers to “ultimate physics”, which is a science perhaps forever beyond human reach, but on the other hand he claims that “physical concepts... are tied to the \textit{common sense notion} of a physical object....”\textsuperscript{941} It seems to me that whenever the mental is contrasted with the physical, it is the commonsense notion of a physical object that is used, which is the opposite of the mental.

Davidson’s reference to the commonsense notion of a physical object is based on his “naturalistic view” that, because humans seem to be inducers by nature, we cannot but observe physical events and objects as being law-governed. This perhaps sounds like a

\textsuperscript{938} Davidson, 1995c, 214, 215. See also Davidson, 1985.
\textsuperscript{939} Davidson, 1995c, 202.
\textsuperscript{940} Davidson, 1985, 227.
\textsuperscript{941} Davidson, 1964, 45, my emphasis.
philosopher’s psychological speculation. But for clarification we can recall Ramberg’s summarization of Davidson’s position. First he notes that, according to Davidson:

> What we count as an object and what we count as a state of an object, as well as what we count as a change, is governed by our fundamental interest in construing our environment in terms of generalities. This is an interest that we have...as biological creatures. When we take two events as causally related, we thereby take them as nomologically related, because to recognize a change in the state of a physical object just is to recognize an event which is susceptible to explanation in terms of empirical law.\(^{942}\)

The “fundamental interest” is obviously not an interest of which humans are aware of at birth. Perhaps it could be claimed that this is what makes it “fundamental”. The interest is part of human biology. Ramberg concludes:

> [...] the observer of physical events cannot but see them as, on the whole, instances of how things generally tend to go. We couldn’t fail to discover general relations by which we understand the changes we perceive in the physical objects...because we are by nature disposed to count as changes and as persistent subjects of such changes whatever will yield general patterns allowing us to predict our environment.\(^{943}\)

The human observer has no alternative but to see reality in a certain way. The nature of physical concepts is what it is because we make predictions by tying the concepts of change and object together. Reality could look very different from the perspective of a creature with a different biological nature. As a consequence also, its PNCC would differ from ours.

The implications of human biological nature for the PNCC become clear when Davidson notes: “[...] the ground floor connection of causality with regularity is not made by experience, but is built into the idea of objects whose changes are causally tied to other changes.”\(^{944}\) But what does it mean to say that the connection between regularity and causality is “built into the idea of objects”? My claim is that we are talking of concepts, and since they are our creations we decide how they are used. But this cannot be wholly our decision because the fact that humans are the kind of creatures they are must be taken into account. In this sense, so it seems to me, the connection between regularity and causality is partly made by experience. If all humans were blind, our visual concepts would differ from the ones that we have now. If all humans were deaf, our auditory concepts would differ from the current ones and so on. That singular causal statements imply the existence of covering laws “is not an

---

\(^{942}\) Ramberg, 1999, 611, my emphasis.

\(^{943}\) Ibid.

\(^{944}\) Davidson, 1985, 227.
empirical fact: nature doesn’t care what we call a change, so we decide what counts as a change on the basis of what we want to explain…”

It is interesting that Davidson attempts to explain an *a priori* view about the nature of causation by referring to human biological nature. Some commentators, like Sinclair, have concluded that Davidson’s constitutive principles are susceptible to empirical revision since they are responsive to empirical features of human beings as biological creatures. A broader conclusion is that “our view of ourselves as agents cannot be separated from important empirical features concerning the type of creatures we are.” Mental phenomena are shown to be *natural* through an account which aims to demonstrate how these states are possible for humans. This kind of interpretation of Davidson’s intentions aims to show that his overall position is sufficiently naturalistic without being reductionist. Sinclair suggests that Davidson can be seen as contributing to a distinctive conception of philosophical naturalism. I agree that this is one plausible interpretation of Davidson’s general philosophical position, but I nevertheless suggest that Davidson’s philosophy is best understood as an attempt to clarify conceptual structures and *not* as an attempt to come up with a distinctive version of naturalism. The purpose is not to offer explanations or theories about the mind. This being said, as I suggested in chapters one and two, it should be emphasized that the non-reductive positions of Davidson or von Wright need not be viewed as forms of *antinaturalism*. Sinclair’s interpretation can be seen as defending this kind of suggestion. In section 1.2.2 I claimed that Davidson’s understanding of the nature, purpose and aims of philosophy is not entirely clear; I believe this is the conclusion that we have to be satisfied with.

One way to understand the connection between laws and causes is the “naturalistic” interpretation, which claims that the truth of the PNCC is based on certain empirical features of humans. But on the other hand, Davidson appeals to the intuition that since concepts are ours it is, in a sense, up to us how they are correctly used. The conviction that the changes of a physical object are governed by laws can be seen as *a priori* if we think that this is part of what we *mean* by our concept of physical objects. I think it is questionable whether our common sense notion actually corresponds with the intuition that Davidson is suggesting; perhaps it does. This is an empirical question. However, we *seem* to treat physical objects as

945 Davidson, 1995c, 212, my emphases.
if they are governed by laws right from the beginning. Human behavior shows that we usually expect physical objects to behave in a lawful manner. This is perhaps the best evidence for the view that our concept of a physical object is a concept of a law-governed object. Perhaps the concept reflects something which comes naturally to humans, namely the capability of detecting certain kind of regularities. Infants expect that a moving object which disappears behind a wall will come out from the other side. If it doesn’t, they seem surprised. If an adult sees a rock flying towards his head at a certain speed he expects it to hit him soon. The examples of how we treat or see physical reality as being law-governed are endless. This suggests, as Ramberg notes, that in a sense we cannot see physical objects and events but as instances of how things generally tend to go, and our concept of a physical object is thus accordingly formulated. How we perceive reality is not up to us, although our interpretation of reality includes also an element of choice.

With respect to the strict laws of physics, it must be stressed that the aim of physics is to provide laws that are as complete and precise as possible. But this is an intrinsic requirement of physics. According to Davidson the “physical vocabulary” of physics, not the common sense physical vocabulary but the scientific one, is committed to the idea that a physical object is an object the changes of which are governed by strict laws because there exists a vocabulary, the vocabulary of physics, into which the broader physical vocabulary could be assimilated. It is unclear whether this suggestion is confirmed by current scientific knowledge. Davidson’s claim that a causal relation can always be described so that cause and effect fall under a strict law cannot be based on evidence; it cannot be empirically confirmed. I think that the PNCC must remain as an assumption because the conceptual argument for it cannot show that causal relations in nature are really backed up by strict laws. It could be claimed that the vocabulary of physics is a vocabulary in which the generality of laws is manifested in a way that side-steps all other concerns except the special pursuit of generality. This reasoning is, however, somewhat circular because, according to Davidson, basic physics is the only field containing a vocabulary sufficient for expressing strict laws. Fundamental physics is characterized as science which aims to find a vocabulary that is complete in the sense that whatever can be described in that vocabulary can also be explained in that vocabulary.

Davidson’s view is that it is “plausible” that there is a system of concepts in terms of which a closed causal system can be formulated. Within such a system of concepts, it is possible to formulate homonomic generalizations the positive instances of which give a reason to believe that these generalizations can be improved by adding further conditions in
the same vocabulary used in the original generalizations. As the term “homonomic” suggests, there is something similar in the concepts of such a system. Davidson notes: “Nomological statements bring together predicates that we know a priori are made for each other....” Here again it is claimed that we can know this a priori. It is difficult to say to what kind of “nomological statements” Davidson is referring, and how we know that a statement is nomological. Is he talking about heteronomic or homonomic generalizations or about strict laws? Each one of these could be taken to be nomological statement in some sense of the word. Lawlikeness is, after all, a matter of degree, as Davidson reminds us. However, I believe we can interpret this view as implying that truly nomological statements equal strict laws, and to qualify as a strict law a statement must satisfy the definition of a strict law, which was given earlier. It remains to be seen what kind of predicates are made for each other in a way that makes them suitable for use in strict laws.

As noted in section 3.1, it is not only in physics but in all the “physical sciences” that homonomic generalizations are possible. Physical sciences hold the promise that their generalizations can be sharpened indefinitely by drawing on further physical concepts. In the physical sciences, and more precisely in physics, there is a “theoretical asymptote of perfect coherence with all the evidence, perfect predictability (under the terms of the system), total explanation (again under the terms of the system).” It is questionable how realistic this view is; the idea of perfect predictability and total explanation may have to be rejected at the quantum level. The view nevertheless expresses Davidson’s understanding of the nature of the physical sciences. It seems that on Davidson’s view, strict laws can be found only in closed theories, that is, in theories where the events in its domain interact only with other events within the theory’s domain. He notes: “[a] law can hope to be precise, explicit, and as exceptionless as possible only if its draws its concepts from a comprehensive closed theory.” A strict law can be formulated only in terms of concepts that are drawn from a theory which treats reality as a closed system. Presumably only physics, because of its aims, is such a theory. This suggests that strict laws can be found only in physics. Could there be other closed theories than physics, say chemistry? If chemistry is a closed theory wouldn’t also its laws qualify as strict? Davidson’s view seems to be that only physics counts as a closed theory because in all other vocabularies, unexplainable events – events that cannot be explained inside the given vocabulary – are possible. Non-chemical events may influence

---

950 Davidson, 1970, 218.
952 Ibid.
chemical events in ways that cannot be described or explained *inside chemistry*. Therefore, chemistry is not a closed system. Physics, on the other hand, treats everything as a possible cause if it falls within its reach. Whatever could influence physical events would count as physical from the perspective of physics.

Davidson’s formulation from “Mental Events” is: “Physical theory promises to provide a comprehensive closed system guaranteed to yield a standardized unique description of every physical event couched in a vocabulary amenable to law.”  

What does it mean to say that physical theory “promises” something? Presumably, that the *purpose or aim* of a physical theory is the one described above. Perhaps this task is forever beyond human possibilities, but we can still aim for it. Needless to say, the success of this project must depend also on the facts about the kind of universe our universe happens to be, on how strictly law-governed the universe actually is. Davidson acknowledges this. He quickly dropped the requirement that strict laws should be thought of as deterministic and claimed that they should be “as deterministic as nature can be found to be”. It is possible that reality may turn out to be hopelessly non-deterministic. I think it could be claimed that the existence of mental events already shows that parts of reality are incurably non-deterministic. If free will can exercise its powers in ways which cannot be predicted, what does it mean to claim that reality is deterministic? It is instructive to recall von Wright’s comment:

> *Determinism holds good, one could say, to the extent that it works i.e. we can successfully predict the future on the basis of past experiences and hypothetically assumed laws of the nature.... In many areas of science the idea is likely to continue to play its classical role. In other areas it may have to be modified ('relaxed') or it will be dropped as useless.*

Is the pursuit of “Davidsonian determinism” a rational task? The concepts of “physical object” and “law” are of our own making, but their applicability depends also on the nature of reality. As von Wright notes, in “those parts of reality” where determinism seems to fail, it would be perhaps better to drop the whole idea. I think that we cannot *really make sense* of the view that our behavior could be strictly predicted or explained. The idea of determinism can work as a guiding principle, as a “heuristic device” for science, but it does not have a clear meaning in the realm of human behavior. Science attempts to transcend our common-

---

953 Davidson, 1970, 224.
954 See Davidson, 1970, 208 and Davidson, 1993d, 191.
955 Von Wright, 1985, 42. See also von Wright, 1998, 150.
sense notions and proceeds to describe the reality “as it is” not “as it shows itself to us”. But this pursuit also has its limits. Davidson notes:

[…] it is hard to think that the question whether such [a completely deterministic theory] exists is a purely conceptual question, at least a theory human beings could, even in principle, invent and test. Surely we must allow that the best physics that is possible for us is irreducibly probabilistic…if physics cannot be made deterministic, if the ultimate laws of the universe, so far as we will ever know, are probabilistic, then we must think of causality as probabilistic. Singular causal statements will still entail the existence of strict laws… but the laws will not meet Hume’s or Kant’s or Einstein’s standards.  

The PNCC still holds, but the verdict on the degree of nomologicality secured under it will be set by our finished physics. It is clear that in Davidson’s view strict laws are to be found only inside physics: “I made clear that what I was calling a law in this context [Mental Events] was something that one could at best hope to find in a developed physics…. I allowed that there are not, and perhaps could not be expected to be, laws of this sort in the special sciences.” Most critics, for example, Kim did not read the original message of “Mental Events” in this way. Davidson’s admittance that strict laws are at best found in a developed physics. It sounds like the definition of a strict law is such that only the laws of physics will qualify as strict. Given this view, it is strange that Davidson continues to say that there are no such laws in the special sciences. The assertion that there could not be expected to be such laws in the special sciences is puzzling. If, by definition, strict laws are only to be found in physics (because of its aim of closure) it seems to follow immediately that these laws are not to be found elsewhere.

It seems to follow from Davidson’s admittance that homonomic laws are not the same thing as strict laws. Strict laws should not be confused with other regularities. In his insightful article discussing the anomalism of the mental, Brian McLaughlin has claimed: “A strict law (or expression) is a homonomic law (or law expression)…. The distinction between homonomic and heteronomic laws is just the distinction between strict and unstrict laws.” When the argument for the anomalism of the mental was first introduced, Davidson operated with the concepts “heteronomic” and “homonomic” generalizations”. It seems to me that the concept of a strict law must be added to this distinction in order to understand Davidson’s argument correctly. If this is true, two questions can be raised with respect to McLaughlin’s interpretation. First, how does it fit with Davidson’s claim that there are “[...] rough, but homonomic, laws....”? The second question is simply whether McLaughlin’s interpretation is correct. As far as the second question is concerned, McLaughlin’s view depends on his

---

957 Davidson, 1993d, 191.
958 McLaughlin, 1985, 342.
interpretation that: “A law is strict when but only when it is couched solely in the extended basic vocabulary of a closed comprehensive theory.” Mclaughlin refers to an extended basic vocabulary, which may contain terms which are linked to the basic vocabulary via bridge laws. On the other hand, he mentions that a more restricted version of a strict law could be defined. According to this definition, a law is strict if formulated solely in the basic vocabulary of a closed theory. Mclaughlin assumes that in this sense only, the laws of physics would count as strict laws. He however concludes that his liberal definition better fits Davidson’s use of the term. With respect to the first question it is difficult to accept the view that, given Davidson’s definition, a strict law could be “rough.” Yet, since Mclaughlin identifies strict laws with homonomic laws, he should interpret Davidson as suggesting that rough homonomic laws are the same thing as rough strict laws. But what could this mean?

Looking back to “Mental Events”, it could be claimed that Mclaughlin’s interpretation perhaps fits better with Davidson’s use of the concept of a strict law than the interpretation according to which strict laws must be couched solely in the basic vocabulary of a closed theory. However, if we look to “Laws and Cause”, which is partly a refinement of “Mental Events”, it seems evident that Mclaughlin cannot be correct. Strict laws cannot be “rough” and they must be couched in the vocabulary of physics. This being said, I believe that we must still somehow squeeze out an interpretation which in essential respect resembles the one given by Mclaughlin. Why? Because by taking seriously the stronger interpretation, we encounter a problem which may turn out to be fatal for the argument against the possibility of strict psycho-physical laws. Since I think that the argument against psycho-physical laws is, by and large, interesting, I try to clear out the problem which the strong interpretation creates, and thereby to defend the consistency of Davidson’s views. The seriousness of the problem depends, on the one hand, on our understanding of the notion of a strict law; on the other, it also depends on our understanding of the notion of a physical science.

An essential question is that of which predicates are suitable for strict laws. Only predicates suitable for strict laws can be reduced to each other. It seems that, if we listen to Mclaughlin, the possible predicates must be either the predicates of physics or predicates which are reducible to these predicates. However, by definition, strict laws are only to be found in developed physics, so how could there be a possibility that predicates from other “physical sciences” could be reduced to these “strict” predicates via “strict bridging laws”? There seems to be no possibility for such a reduction. If this is the case, then Kim’s

---

960 Mclaughlin, 1985, 346.
observation that reductions in the sciences are impossible if Davidson’s view is accepted must be taken seriously. Kim’s conclusion is:

[...] if we are allowed to begin with the nonexistence of strict laws outside basic physics as a premise, we can cut away all the complex and obscure arguments... for mental anomalism and reach the desired conclusion in a single step. For since strict laws can exist only in basic physics, there of course cannot be strict laws in psychology or between psychology and anything else.\textsuperscript{961}

This is a very interesting observation and, if correct, it makes one wonder why Davidson wants to provide arguments against strict psycho-physical laws if, by stipulation there cannot be strict laws anywhere outside physics. Moreover, if there are no strict bridging laws between physics and anything else, our mental vocabulary does not deserve special status when compared to any other special science vocabulary, which also turns out to be irreducible. The mental would not be a special category in any way.\textsuperscript{962} It seems to me that Kim’s interpretation must be rejected if the thesis about the lawlessness of the mental is to remain interesting. This, however, is very difficult because Davidson explicitly admits that strict laws exist only in a developed physics. I think that we should try to find a way to make room for strict bridging laws between physics and other domains. Mclaughlin’s interpretation should be defended against Kim’s, although given Davidson’s own comments the latter sounds more plausible. Another way to argue against Kim is to explore the possibility of non-strict bridging laws.

Let us for the moment ignore Davidson’s views about the connection between strict laws and physics. In the article where this connection is explicitly made, Davidson also notes that the sign of a heteronomic generalization is that it is not in the form of a strict law and not reducible to such.\textsuperscript{963} This suggests that some generalizations, presumably homonomic, either are strict laws or are reducible to such laws. Indeed, when discussing “rough homonomic laws”, Davidson argues that there are laws which “back the homonomic laws up” and these laws are such that there is no improving on them in points of precision and comprehensiveness. It seems that these are the “real” strict laws, because they are such that they cannot be sharpened anymore. This is of course one reason to reject Mclaughlin’s suggestion that homonomic laws are strict, because it is not the case that every homonomic law is a strict law. The former can be improved, the latter not. The important question is whether homonomic generalizations can be nomologically reduced to strict laws? If they can

\textsuperscript{961} Kim, 1998, 93 fn. 6. See also Kim, 1993.
\textsuperscript{962} This is Rorty’s (1998) reading of the irreducibility thesis.
\textsuperscript{963} Davidson, 1993.
be so reduced, certain domains could be reduced to the vocabulary of physics. If they cannot
be reduced, it would seem that in this respect homonomic generalizations would stand on the
same level as heteronomic generalizations. This being the case, one could further argue that,
in the end, the case of psycho-physical laws is no worse than the case of physical-physics
laws. The irreducibility of the mental would thus not amount to a very interesting claim. But
if rough homonomic generalizations are reducible to strict laws, then there must be bridge
laws which are not strict and which make the reduction between domains possible. One could
think of homonomic generalizations as being laws of some natural science. Such a law of
domain S (the science in question) cannot be reduced to the laws of physics via strict laws
because strict laws exist only inside physics. Yet, homonomic generalizations can be
“improved”, which I take to mean that they could be sharpened, in theory at least, into strict
laws.

Saying that in the questions of reducibility mental and physical vocabularies are not
very different or far apart is not how most commentators have interpreted Davidson’s
position. A standard interpretation of Davidson’s view is that our physical and mental
vocabularies are essentially different. This kind of view is strongly suggested, for example, in
Davidson’s early article “On Mental Concepts and Physical Concepts”. Events falling under
our mental vocabulary are not predictable or explainable by strict deterministic laws, whereas
events falling under our physical vocabulary are. The following captures the general intuition
that I take to have motivated “Mental Events”. Davidson asks: “Can intentional human
behavior be explained and predicted in the same way other phenomena are?” The part of
the answer which is relevant to making the point that I am pressing is that “there are good
arguments against the view that thought, desire and voluntary action can be brought under
deterministic laws, as physical phenomena can.” The claim is that explanations in terms of
mental concepts are different than all other explanations. This being the case, it is tempting to
think that there must be a sense in which our mental vocabulary cannot ever reduce to physics
in any relevant sense, whereas other vocabularies do, or may, reduce. It is this assumption
which leads to the statement that explanatory: “[...] accounts of intentional behaviour operate
in a conceptual framework removed from the direct reach of physical law....” or more
generally that “mental events... resist capture in the nomological net of physical theory.”
We should consider how strange it would be to say that physical events resist capture in the

---

964 Davidson, 1974b, 230, my emphasis.
965 Ibid. My emphasis.
966 Davidson, 1970, 225.
967 Davidson, 1970, 207.
nomological net of physical theory. A physical event should surely be explainable and predictable inside its nomological net, and this net covers everything.

It seems obvious that a strong and principled distinction between mental events and physical events must be drawn. Otherwise we would have to state something as absurd as: mental events, as well as physical events, resist capture in the nomological net of a physical theory. I think we must find out whether Davidson makes a principled distinction between our mental and physical vocabularies. If he does make it, on what is it based? I think we should also explicate the sense in which physical events are reducible to physics, if they indeed are. The way in which they are reducible must be such that this kind of reduction is impossible in the case of mental events. Kim argues that since strict laws can be found only inside physics, there cannot be reductions via strict laws to physics. It seems that if Davidson is to allow the in-principle reduction of physical events, such a reduction cannot require strict laws.

Whether a principled distinction between the mental and the physical can be made can be clarified by considering whether Davidson draws a line only between the “physical sciences” and psychology, or whether he distinguishes physics as a separate domain from which all other sciences should be separated because of their irreducibility. If he draws only the former line, there is some reason to argue that sciences or vocabularies using mental concepts should be separated from vocabularies that do not use these concepts. The latter as well as the former make up autonomous domains with their own constitutive principles, laws and concepts. If, on the other hand, Davidson separates reality into more conceptual domains than two, physics being one of them, it is not clear what the relations between them are. Most importantly, the relation between the mental and the rest becomes obscure because there seems to be no special reason to emphasize the irreducibility of the mental if other vocabularies are also irreducible to physics. I think that if it turns out that, say, geology is as irreducible to physics as psychology, we could nevertheless ask why the mental vocabulary is irreducible and see whether something makes the irreducibility of the mental a special case after all. The question would be whether the irreducibility of the mental differs from the irreducibility of the geological.

According to Davidson the anomalism or irreducibility of the mental is a fact which is known a priori. Remarks that confirm this are found all over his writings. Our mental vocabulary, and therefore psychology which uses mental concepts, cannot be reduced to physics, natural sciences or to our physical vocabulary. This is to say that our common-sense understanding of ourselves, being formulated in the mental vocabulary, is irreducible. Davidson uses terms like “physics”, “natural sciences”, “physical sciences”, “physical
vocabulary”, “physical laws”, “physical terms” or “physical events”, often in a quite an overlapping and sometimes in an inconsistent manner. For example, in a statement like: “A particular physical event, state or disposition is one that can be picked out… using a vocabulary drawn from some physical science”<sup>968</sup> we are not told what counts as a physical science. However, the domain of the physical sciences can be quite clearly defined, and we can assume that Davidson refers to these sciences by the term “physical sciences”<sup>969</sup>.

As long as we stay within the vocabulary of the various physical sciences, it is possible to ask whether there is a reason for the irreducibility of the special sciences, which do not use mental vocabulary. There are few remarks which suggest that, according to Davidson, such a reduction is <i>not</i> possible. He notes, for example, that: “The best descriptions we are able to give of most events are not descriptions that fall under, <i>or will ever fall under</i>, strict laws.”<sup>970</sup> This is a very strong view. Our best descriptions of <i>most</i> events <i>will never</i> fall under strict laws. But what can the expression “best description” mean in this context? Let us consider an event E, which is a physical event, a chemical event, a behavioral event and a psychological event. Which one of these descriptions is the “best”? As I have shown, Davidson thinks that one explanation cannot be better than another outside the context in which it is offered. Shouldn’t the same be said about descriptions? If the answer is affirmative, then it makes no sense to say that the <i>best</i> descriptions of most events cannot fall under strict laws. If the answer is negative, then it should be asked what makes one description better than another, and what the best descriptions of events are. A plausible – but not very convincing and not easily defended – suggestion is that, according to Davidson, the best description of an event would be the description of it given in the language of ideal physics. But if this is the case, then the “best description” of an event <i>falls</i> under a strict law.

The suggestion above surely gives the impression that most vocabularies in which events are described are such that they do not involve strict laws, and more importantly, cannot be reduced to such laws. Indeed, as Davidson notes: “I suppose most of our practical lore (and science) is heteronomic.”<sup>971</sup> When Davidson mentions the prospects of reduction in the special sciences, his general tone is surprisingly skeptical: “The assumption that biology and neurobiology are reducible to physics... is probably false”<sup>972</sup> or “there may be a

---

<sup>968</sup> Davidson, 1990b, 18.
<sup>969</sup> They include at least astronomy, chemistry, physics and many of the earth sciences like geology and meteorology.
<sup>970</sup> Davidson, 1993d, 191, my emphasis.
<sup>971</sup> Davidson, 1970, 219.
<sup>972</sup> Davidson, 1973, 246.
significant sense in which geology, etc., cannot be reduced to physics”⁹⁷³ and “aerodynamics and biology... almost certainly will not be reduced in any strong sense to basic physics.”⁹⁷⁴ These kinds of remarks, especially if they are based on the view that reduction is not forthcoming because the relevant events are not describable in a vocabulary which falls under strict laws, give rise to Kim’s criticism. The criticism, which challenged the need for the arguments for the anomalism of the mental, has also another component – namely, that there isn’t going to be any reduction anywhere in science. This suggestion is obviously false, because there clearly have been reductions in science; Kim’s argument is thus meant to show that Davidson’s view of reducibility is unrealistic.

I think it is important to note that Davidson does not give any clear reasons or arguments why we should be so pessimistic in the case of our non-mental vocabularies. This being the case, there is not much point in trying to speculate about what Davidson means by his remarks that the assumption of reduction is “probably false”, or that certain sciences “almost certainly” will not be reduced to physics “in any strong sense”. It seems to me that the suggestions are mere empirical speculation or Davidson’s personal opinions, ones which are not defended by arguments. The irreducibility of neurobiology would certainly be challenged by reductivists, although they might agree that neurobiology will not reduce to physics. From the perspective of neurobiology this is an irrelevant or uninteresting claim. So what if neurobiology does not reduce to physics? Davidson on the other hand – unlike the scientist – is interested in the principled obstacles to an ultimate reduction to physics, which is the ultimate science.

That Davidson does not know the reasons for the supposed irreducibility of the special sciences becomes evident when he comments: “The laws of many physical sciences are also not like the laws of physics, but I do not know of important theoretical... reasons they cannot be reduced to the laws of physics.”⁹⁷⁵ Here a critic like Kim could reply: “There is a theoretical reason, namely, if bridge laws are strict laws and there are no strict laws outside physics, then there cannot be reduction outside physics.” But, contrary to the above, Davidson seems to argue that there could be reductions outside physics. An example clearly showing this is the following:

I suppose it possible that there are no meteorological or geological events that cannot in theory be explained by the physics we know; in theory we could describe these events...in the language of pure

⁹⁷³ Davidson, 1987a, 114.
⁹⁷⁴ Davidson, 1999a, 620.
⁹⁷⁵ Davidson, 1987a, 112.
It is unclear how this claim fits in with the views described earlier. The passage nevertheless clearly shows Davidson’s views about the theoretical possibility of reductions between vocabularies that belong to the same conceptual domain. Here Davidson suggests that a reduction of the special sciences to the language of physics is possible, at least in theory. This kind of reducibility is impossible in the case of the mental. Mental concepts cannot be formulated in the language of physics and, this being the case, the laws of physics cannot be used to explain and predict mental events, whereas geological events, for example, might be so predicted. “Geological determinism” but not psychological determinism, is a possibility.

In some sense, the physical and natural sciences belong to the same conceptual domain. Since psychological descriptions are not reducible to this, i.e. to the whole of our physical vocabulary, they are not suited to incorporation into any strict system of laws. On occasion Davidson uses the very expression “strict system of laws”; this raises the question of whether it is only meant to denote the system of strict laws of physics or also some other systems that have laws that could be reduced to the laws of physics. Davidson’s view suggests that there is a general scheme, our physical vocabulary including everything but the mental, which is nomological and to which the mental does not reduce. Reduction between nomological schemes like physics, neurology and biology is possible, at least in principle. Why? Because “we think that though the concepts of biology may not be definable in the concepts of physics, the phenomena of which biology treats can be understood as belonging to the same conceptual domain as that of an inclusive physics.” More precisely: “one can hope for strict connecting laws... when the concepts connected by the laws are based on the criteria of the same sort....” It is thus suggested that strict connecting laws – bridge laws – can exist only in cases where the concepts in the laws are based on criteria of the “same sort.”

But what concepts are of the same sort? Davidson does not give a clear answer, but in light of what has been said, I conclude that it is plausible to think that our whole “physical vocabulary”, which includes everything except mental phenomena, is based on the same sort of criteria. Indeed, the definition of a homonomic generalization from “Mental Events” suggests that these generalizations are such that they can be improved by drawing on the

\[\text{\footnotesize 976 Davidson, 1993e, 312.}\]
\[\text{\footnotesize 977 This is the way that many read him. Stoutland (2005) notes that for Davidson the term "physical" usually denotes the nomological.}\]
\[\text{\footnotesize 978 Davidson, 1999n, 124, my emphasis.}\]
\[\text{\footnotesize 979 Davidson, 1991, 162.}\]
vocabularies from the same field as the original vocabulary in which the generalization was stated. In this process, the generalization could, in principle, be turned into a strict law which would be a law of our developed physics.

The concepts of physics are all based on the same sort of criteria; but I take Davidson to be suggesting that the principles governing the use of a broader physical vocabulary are in some sense similar to those of physics. The interpretation that all physical properties or predicates are based on the same kind of criteria is supported by Davidson’s admittance that:

In the case of causal properties like elasticity, slipperiness, malleability, or solubility we tend to think, right or wrongly, that what they leave unexplained can be (or already has been) explained by the advance of science. We would not be changing the subject if we were to drop the concept of elasticity in favour of a specification of the microstructure of the materials in the airplane wing that cause it to return to its original shape when exposed to certain forces.\(^\text{980}\)

This suggests that generalizations involving concepts like “solubility” could be sharpened to a point where the relevant laws could not be improved any more. Because a physical concept like solubility belongs to the same conceptual domain as the concepts of physics, this concept could be replaced with a description of a mechanism or ultimately a microstructure which explains why something is soluble or what solubility is. Nothing that is relevant for physical explanation would be lost if a detailed description of microstructure, instead of the term “solubility”, were to be used. On the contrary, the explanation that refers to the microstructure would explain why a sample S dissolves whereas sample S\(_1\) does not, although both samples seem relatively similar at the macrolevel. An explanation that one sample is soluble and the other is not merely indicates that there is something which explains the difference; an explanation not referring to solubility would explain this problem away.

But when and why do events fall under the same conceptual domain? If we consider the view given earlier, namely: “It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and conditions physically described”,\(^\text{981}\) then we get an answer to the question of why we should think that the physical vocabulary constitutes a homogeneous conceptual domain. Although there is conceptual diversity between the various physical vocabularies we use, in the sense that they use very different concepts and function at very different levels, there is nevertheless an idea of explanatory unity. Why? Because our concept of a physical object or event is allegedly such that its behavior is law-governed, and every physical science studying these objects and events

\(^{980}\) Davidson, 1991, 163.

\(^{981}\) Davidson, 1970, 222.
attempts to formulate laws in terms of which physical explanations could be given. It could be argued that all physical characterizations, as long as they are meant to be predictive, are meant to describe the general patterns of changes that their respective objects undergo. The articulation of these general patterns amounts to providing a body of laws or lawlike generalizations. When this idea is further generalized, it can be suggested that there exists, at least in theory, a closed and deterministic system of laws into which physical events fit. So, there are our physical vocabularies which are “law-governed”, and there is our mental vocabulary which is governed by normative principles. These two domains are conceptually too different for strong connections to exist between them.

It is quite obvious that “the mental” should be contrasted with the physical, not with physics. Explanations in physics, insofar as they can be called explanations, answer to the highest standards of nomologicality that can be imagined. It is therefore philosophically instructive to ask whether psycho-physical laws could ever be as precise as the strictest of known laws. The main interest of this question is theoretical or perhaps we should say, philosophical. I suggest, however, that because of questions relating to human freedom we are interested in knowing whether mental events could somehow elude prediction in terms of the strictest laws available. This is a pragmatic question. The answer to the question of whether psycho-physical laws could be strict would be affirmative if these laws would themselves be strict laws or if they could be reduced to such laws. Davidson denies both possibilities. The case of “non-mental” sciences is different. Although the laws of these sciences are not strict, their vocabularies could, in principle, be reduced to physics. Although Davidson is not always clear on this question, it seems nevertheless to be the case that the “laws” governing mental–physical or mental–mental relations are different in kind than the laws involving purely physical events. It seems to me that Davidson’s notion of a “physical law” is meant to refer to laws between physical events and not to the laws of physics. Physical laws are such that they can, in theory, be sharpened into strict laws, whereas this is not possible for psycho-physical or psychological laws.

The mental should be contrasted with the physical and an additional physical – physics distinction based on the irreducibility of the former should not be drawn. However there remains the question, suggested by Kim, of how the reduction of the special sciences

---

982 As Davidson (1987a, 111) says: ”Since I was interested in the question whether reason-explanations are or ever could be just like the best explanations for which physics strives, I set very high standards for what I called ‘strict’ laws….“ Most commentators read the message of “Mental Events” differently. They thought that the claim was that there cannot be any kind of laws between the mental and the physical. Davidson’s redefinition makes the original argument much weaker than what critics thought it to be.
could be possible, even in principle, if there cannot be strict bridge laws outside physics. Davidson notes that the anomalism of the mental rules out psycho-physical reduction “by way of strict bridging laws....” But if it is the absence of strict psycho-physical laws which rules out this kind of reduction, the absence of strict “physical–physics” bridge laws should rule out the possibility of reduction in the special sciences as well. However, Davidson seems to claim that bridge laws between physical events are possible. Although Davidson talks about strict bridge-laws, it is by no means clear that the term “strict” should be understood in the same sense as in the claim that strict laws are to be found only in physics. Davidson’s use of inaccurate terminology is perhaps thus a very simple solution to Kim’s problem.

I think we should conclude that in Davidson’s terminology, the term strict law is meant to refer just to the laws of physics. It is not surprising that only the laws of physics are strict, since their strictness could be taken to be a definitional feature of them – i.e. the best laws of physics are, by definition, strict. Whether those kinds of laws can ever be found is an open question, but since these laws are generalizations which are “as deterministic as nature can be found to be”, it seems that what counts as a strict law depends, in some sense, on our opinion concerning whether or not we have reached a sufficient level of determinism. It is, of course, also possible to claim that the most basic laws of physics count as “strict laws” no matter what their strictness. On the other hand, as we saw in section 3.1, Davidson has explained in detail the nature of strict laws. But it is necessary to understand on what this view is based. Is the exact definition of strict law just Davidson’s invention?

Davidson claims that physics “sets out” or “strives” to provide exceptionless laws. This being the case, the crucial question is whether certain phenomena can be reduced to the concepts that physics uses. Since, according to Davidson, every event is an event that can be described in the vocabulary of physics and every causal relation can be described in the vocabulary of physics, on a Davidsonian account it is possible to dream of a “full coverage” where the behavior of every physical event and object could be explained and predicted in the vocabulary of physics. Perhaps this is an idle dream, but if in the spirit of O-physicalism it is supposed that reality is thoroughly physical, what does this mean other than an expression of the conviction that the nature of this reality could be completely captured by physics, which has this very task as its aim. Since all events and objects are physical, the question of whether they can be candidates for strict laws depends, partly, on whether they can be identified in the vocabulary of physics. I believe that with respect to this question, the issue of the

---

983 Davidson, 1994a, 232.
irreducibility of the mental is extremely relevant. However, the original argument against strict psychophysical laws seems not to depend on this, at least not entirely.

Let us now turn to a discussion of what prevents the formulation of strict laws between mental and physical phenomena. The original thesis of the anomalism of the mental was that there are no strict deterministic laws on the basis of which mental events can be predicted and explained. As I have noted, Davidson later claimed that at the time of the original statement he thought that strict deterministic laws could exist only in a developed physics. The original claim should therefore be understood as suggesting that mental-physical laws cannot ever be as precise as the laws inside the most developed physics. Perhaps this is indeed what Davidson intended already in 1970. But this is not how most critics have interpreted him. 984 Their interpretation of Davidson’s claim was that bridge laws between the mental and the physical are impossible. They thought that in the argument against type-identity, the expression “bridge law” does not have to be understood as strictly as Davidson intends. Thus, Kim asks:

Why insist on reduction by strict laws only? What’s wrong with non-strict psycho-physical laws as ‘bridge’ laws?.... This [reduction via strict laws] surely cannot be a sense of reduction that holds serious philosophical interest for us. If psychology is reducible by the same standards that apply to the best cases of theory reduction in the sciences... why isn’t that reduction enough? 985

This is surely an important and interesting question. Kim questioned whether there is any need for an argument for the anomalism of the mental if it is the case that, by definition, there cannot be the kind of laws that are being denied. Here he asks what Davidson’s argument against non-strict bridge laws is, and whether their existence would show that psychology is as reducible as any other science that reduces via non-strict laws. I have already argued that Davidson sees a principled difference between our mental and physical vocabularies. But, interestingly, Davidson thinks that the existence of non-strict bridge laws would not show that the mental is reducible to the physical. As he claims: “if the [psycho-physical] laws are not strict the threat [of reduction] is averted and the promise [of reduction] is false.” 986 Why non-strict laws do not hold a promise of reduction will be clarified in more detail in later sections.

One possibility is to argue that Davidson’s anomalism of the mental is a result of his acceptance of laws which are too strict and drawn from physics, which for Davidson is the master science. It could be claimed that anomalism looks plausible only if the ideal laws of

984 For discussion see Robinson, 2001.
986 Davidson, 1993d, 192.
physics are accepted as a yardstick, in terms of which mental events cannot be predicted. Davidson does not give a clear answer to the question of what is wrong with non-strict bridge laws, and I think that Kim is right in his observation that the possibility of reduction through non-strict laws “is not an idle question; nor is it merely a verbal issue.” What seems to be also correct is Kim’s observation that “non-strict laws’ are bad news for anomalous monists. In embracing them we may end up losing anomalism from anomalous monism.” Indeed, Davidson himself has noted that his position is not, in essential respects, different from the position of Fodor, who defends a token-identity theory and claims that there are intentional causal laws which are not strict. Davidson notes: “In the sense in which Kim and Fodor think there are laws linking mental and physical concepts, I also think there are laws….” This important admittance is often neglected by commentators. Kim defends reductionism –as, despite his occasional remarks on the contrary, does Fodor.

That Davidson likens his position to Fodor’s and Kim’s is puzzling. He actually claims that there are endless type–type uniformities or laws on which we depend all the time. The essential claim of AM is only that these uniformities cannot be turned into strict laws of the sort that physics sets out to provide. This means that, according to Davidson, AM is not in conflict with reductive naturalistic theories of the mind but is in fact consistent with them. This is certainly not how most commentators have interpreted the status of AM. Antony, for example, claims that “Davidson categorically denies the possibility of any systematic empirical treatment of psychological states and events.” But this is simply false. For example, in one of his earliest articles Davidson explicitly stated: “Nothing I have said implies that we cannot give causal scientific explanations of particular human actions, thoughts and the like.” Later he has claimed: “[...] there is no difficulty in general in explaining mental events by appeal to neurophysiological or physical causes....” These are not anti-reductionist remarks. On the other hand, Davidson has claimed that psychology and the social sciences are impossible and that no non-mental science could be expected to explain thinking or the states of belief. It is this official anti-reductionist side of Davidson which leads commentators to the following kind of interpretations: “Davidson has to say that

---

988 Ibid.
989 Davidson, 1993d, 194.
990 See Davidson, 1993d, 2001f. Davidson goes even as far as claiming that, although Fodor does not realize it, Fodor’s position is a defense of AM.
991 Antony, 1989, 175.
992 Antony seems to realize this in a later article. See Antony, 1995.
993 Davidson, 1964, 229.
994 Davidson, 1982a, 180.
there is a point in our investigation of mental phenomena where nothing more can be said which will inform us about how these phenomena are related to phenomena described in the physical vocabulary.”

All this being said, we should note that for Davidson the question of the nature of bridge-laws may be partly terminological. Although the different formulations of the thesis of the anomalism of the mental are – presumably – meant to express the same view, they may cause confusions. I have already discussed what the claim was in 1970 and how it was described later in “Thinking Causes”. It could be claimed that I am just quarrelling about mere terminological matters. I believe, however, that we must be very clear on what is actually being denied. The pressing question is what, if anything, is wrong with non-strict reductive bridge laws. There seems to be something wrong with them, because Davidson claims: “…if the [bridge] laws are not strict, the threat [of reduction] is averted.” But this view sounds obscure if the claim that strict laws are to be found only inside physics is also accepted. If strict laws are found only in physics, then Kim is right that Davidson cannot expect to find reduction to physics anywhere.

Davidson has formulated the thesis of the anomalism of the mental in different ways. He notes: “We do not have and cannot expect to find, a way of mapping events described in the physical vocabulary onto events described in the mental vocabulary.” This view resembles the one given already in 1964: “The network of mental concepts and the network of physical concepts are not like two alternative schemes where we can pass in some systematic way from one scheme to the other. There is no formula for doing this.” This claim emphasizes that these two vocabularies are not alternatives to each other. I stressed this point earlier when I claimed that the vocabularies are different; they are based on different kind of evidence and have differing functions. In the quotes above, it is claimed that there is no “mapping” between the two vocabularies. It is not entirely clear what the term “mapping” means. Does it mean that there cannot be bridge-laws of any kind between the domains? This is a plausible interpretation because the claim is that mental and physical events cannot be related. Davidson writes also: “There are no empirical laws linking mental phenomena to physical phenomena,” and we read that mental concepts are not reducible by “natural law”

---

995 Crane, 2000, 80.
996 Davidson, 1993d, 10.
997 Davidson, 1997a, 127, my emphasis.
998 Davidson, 1964, 48, my emphasis.
999 Davidson, 1999o, 599.
to physical concepts. What is the role of the term “empirical” in this claim? What is an empirical law? Is a natural law a law of physics? Does Davidson claim that empirical laws linking mental and physical phenomena do not exist, or is he claiming that such laws cannot be found? Is he suggesting that one cannot find laws of this sort or that there are no such laws? Is there a real difference between these two options? If Davidson is denying the possibility of strict bridge laws, is he also denying the possibility of non-strict bridge laws? He writes: “Ontological reduction does not imply that there are causal or bridging laws relating events classed by mental properties with event classed by physical properties.”

It is unclear to me what the terms “causal” and “bridging” laws mean, but the impression that there are no bridging laws whatsoever between the mental and the physical is quite natural. Indeed, Davidson’s general conclusion is: “There are no precise bridging laws that firmly and reliably relate mental events to physical events.”

The expressions “precise bridging laws”, “firmly” and “reliably” are open to interpretation and therefore the previous statement is extremely vague. Perhaps Davidson intended the expression “precise bridging laws” as a synonym for “strict bridge laws”. A further complaint is that sometimes Davidson talks about the impossibility of laws which connect properties, and sometimes he talks of the impossibility of laws connecting events. The first claim is curious given Davidson’s view about properties. If there are no such things, there cannot be laws connecting them either. The second claim is curious because, as far as I can see, it is unclear how particular events could be related. The only way to understand the latter claim is to acknowledge that events enter law-relations only by having properties which earn them a membership in a certain type.

I think that at this point we have to settle for the view that Davidson is claiming only that the laws involving the mental and the physical are not as precise as the laws of the most developed physics. This would mean that Davidson’s position would agree in spirit with modestly reductive theories of mind, although the motivation to develop “a theory of mind” differs between Davidson and reductivists. This is a disappointing result, especially since it leaves open Kim’s pressing question about what the problem of non-strict bridging laws is.

Davidson, 1993d.

Davidson, 1985, 244. It should be noted how easily Davidson here talks about properties in terms of which events are classed.

Davidson, 1995a, 11.
3.1.2 Why strict psychophysical laws are impossible: A Concrete example

Let us consider the following kind of bridge-law:

“X is water iff X is H$_2$O”.

From this we can conclude that a sample of water is also sample of H$_2$O. The law makes a claim about identity, but it could also be claimed that the law refers to two realms of concepts; it refers to the ordinary concept “water” and to the chemical concept “H$_2$O”, formulating a bridge-law between the concepts. The established identity is a result of empirical investigation. Von Wright has noted that an identity statement like “water is H$_2$O” is contingent, but if we conclude that chemical composition is the best way to identify water, if chemical structure is the surest criterion, then the identity claim is no longer contingent.\textsuperscript{1003}

The suggestion that the status of the identity claim depends on the question of how to identify something as a sample of water is interesting when applied to the alleged mental-physical identities. It suggests how to think of the status of the alleged identity. After all, philosophers are not unanimous in their views about whether or not identity claims like the one above should be thought as being necessary or contingent.\textsuperscript{1004}

Consider a set of mental concepts (M) and a set of physical concepts (P). We have formulated a bridge law between water, an everyday concept, and H$_2$O, a scientific concept. Why is it impossible to formulate bridge-laws between mental concepts, which are part of our commonsense vocabulary, and physical concepts which are part of our scientific picture of the world? Mental–physical bridge laws would be important for the reductivist because they would enable the reducibility of “theory” T (a common sense vocabulary) to theory T$_1$ (a neuroscientific theory); such reducibility would be shown if the terms of T could be formulated in the language of T$_1$. With bridge laws connecting mental and physical properties, the correctness of reductive physicalism or eliminativism could be shown. Every property would be a physical property and a unity would be reached at the ontological and at the explanatory levels. But why are laws of the form (where “m” refers to a mental state and “p” a physical state) impossible:

\textsuperscript{1003} Von Wright, UPg.

\textsuperscript{1004} For recent discussions, see, for example, Barnett, 2000, Anderson, 2005 and Bird, 2001.
“Necessarily, a person is in m iff a person is in p”

or

“Necessarily, a person is in p iff a person is in m”

or

“Necessarily, whenever a person is in m at time t he will be in p at time t₁”

or

“Necessarily, whenever a person is in p at time t he will be in m at time t₁”

Generally speaking why is it the case that there are no true statements of the form:

“X is m iff x is p”? 

A trivial answer is that certain identity claims are contingent. So the laws or statements described above cannot be necessary. The statement “X is m iff x is p” is only contingently true. However, as von Wright notes, the statement “Water is the same as H₂O” can be thought to express a non-contingent truth if the chemical composition of H₂O is taken to be a criterion of water. Why are the identity statements involving mental and physical terms different than the claims involving just physical terms? Bridge-laws connect physical vocabularies. Laws of succession, relating physical vocabularies, state that if something of a certain type exists at some time, then something of another type exists at a related time. Why are these kinds of laws are impossible when one of the vocabularies is mental?

A mental–physical bridge law connecting M and P could state that whenever a person believes that “snakes are reptiles” he is in a certain physical state, which occurs always when this belief occurs; thus, the states, or properties M and P are identical. It is plausible to think that person’s belief that such and such prevents her from believing something else. If a person believes that A, normative principles of belief-attribution prevent her from believing not-A. If A entails B and a person believes that A, she should not believe that not-B. But let us now suppose that a bridge law connects the belief A with a physical state P. Let us also suppose that another bridge law connects the belief not-A with a physical state R. If we know that someone believes that A, we can infer that he does not believe not-A. Likewise, if A entails B and a person believes that A, she should not believe that not-B. But what would it mean to say

1005 I assume that Davidson’s term “causal laws” refers to these kind of laws.
that we can infer from being in P to not being in R? It could mean that we believe that P and R are physically incompatible. But what would it mean to say that, when events are described purely in physical terms, one event should not have a connection to another?

Physical states of course have certain kind of connections between them. But whatever these causal connections actually are, they cannot be a reason for saying that they should be in a certain way. We can say that given our expectations concerning the results of a specific physical experiment, a certain result should occur. But this is surely not the same way in which we say that a person who believes that “It is raining” should not believe that it is not raining. In the former case the “should” reflects our expectations about the outcome of the experiment, and there are no principled reasons why those expectations could not fail. In the latter case, at least according to interpretationism, the existence of the belief “It is raining” depends on our view that if one has it, then one should not have the belief that it is not raining. The identity of a belief depends on the connections it has to other beliefs. If a person seems to hold beliefs “It is raining” and “It is not raining”, we have to question whether either one of these beliefs can be attributed to her. In the psychological or mental reality, the idea that somebody ought to think something given what else she thinks is a guiding principle. In physical reality certain empirical and logical principles constraint the mode of being of things, but there is no sense in which objects ought to be one way rather than another. The “ought” is not a constitutive principle of the physical. Likewise, empirical constraints are not relevant constitutive principles of the mental. The structure of mental reality is constituted in interpretation, and this process is governed by non-empirical principles. Whether or not something is soluble is an empirical question, but the question of whether or not a thinker subscribes to the basic principles of rationality is not empirical.

The argument against strict psycho-physical laws is certainly difficult to accept for those who do not share Davidson’s view about the necessity of interpretation. The following two statements suggest why strict laws connecting the mental and the physical vocabularies are impossible, and how this is related to the Davidsonian conception of the mental. Davidson notes:

Davidson (1986, 208) notes: "[...] with desire as with belief, there is a presumption… that similar causes beget similar evaluations in interpreter and interpreted. This is not, I should emphasize, either an empirical claim or an assumption for the sake of good science. It is a necessary condition of correct interpretation.” In the physical realm, the assumption that similar causes beget similar effects is an empirical claim or an assumption for the sake of good science. It can be thought of, as von Wright notes, as a scientific norm guiding research, but violations of it are certainly possible. Not so in the case of interpretation.
There are no strict laws because of the disparate commitments of the mental and physical schemes…. There cannot be tight connections between the realms if each is to retain allegiance to its proper source of evidence.\textsuperscript{1007}

Moreover:

[…] to allow the possibility of such laws would amount to changing the subject. By changing the subject I mean here: deciding not to accept the criterion of the mental in terms of the vocabulary of the propositional attitudes.\textsuperscript{1008}

What may sound mysterious is, on the one hand, the claim that different schemes would have disparate commitments and, on the other hand, the claim that tight connections are impossible if each scheme is to remain faithful to its proper source of evidence. Could it not be argued that there is no reason why a scheme should retain allegiance to its “proper source of evidence”? Davidson’s conclusion would be that by allowing this “a change of subject” would occur; the result of such change would be that mental phenomena would not be understood in terms of propositional attitudes.\textsuperscript{1009} Interestingly, Malcolm has noted:

We say we know our friend is crying from pain, because we saw him crack his shin on the porch step. But if we were asked whether we are sure that such-and-such a process is taking place in his brain, we should not understand the relevance of the question: it would be an irritating change of subject.\textsuperscript{1010}

We can note that Malcolm and Davidson wrote about the “change of subject” only a year apart. Their argument can be said to be based on a Wittgensteinian distinction between symptoms and criteria. That Malcolm makes much use of this distinction is clear, but it seems to me that the distinction works as a guiding principle also behind “Mental Events”.

I claim that from the Davidsonian perspective, the real reason against the possibility of psycho-physical laws is thus our unwillingness to give up the vocabulary of folk-psychology – our unwillingness to change the subject. Both Davidson and Malcolm agree that, in some sense, one could decide not to accept the current criteria for the mental as constitutive criteria. I think that this decision, if it were honestly made, should then have wider consequences; brain processes would be the criteria on the basis of which mental phenomena would be attributed. What kind of consequences would this have for human life? Malcolm notes that if somebody did not agree with us, if he decided to use different criteria for the attribution of mental phenomena, then “We could not say that he was wrong… but he would be different.

\textsuperscript{1007} Davidson, 1970, 222.
\textsuperscript{1008} Davidson, 1970, 216.
\textsuperscript{1009} For a warning against the change of subject, see also Davidson, 1974b, 1985.
\textsuperscript{1010} Malcolm, 1971, 63, my emphasis.
He would not have the concepts of pain, of consciousness, or of any mental predicates, \textit{in the same way} we do. I am also inclined to say that he does not have the concept of a person, or at least not the same one we have.\textsuperscript{1011} This kind of view has been subsequently emphasized by Bennett and Hacker. We can talk of brains which make interpretations or decisions, or machines which believe this and that. If we accept this kind of talk, we must realize that the conceptual connections which the concepts of decision or belief have in our current ways of talking are lost when the concepts are used outside their usual field of application. Then it is very questionable whether the terms – superficially similar to our familiar concepts – would be the same concepts at all to which we are used to.

One reason for Davidson’s unwillingness to give up the common-sense criterion for mental phenomena is his conviction that we have no alternative to our folk-psychological interpretation-based understanding of mental phenomena. But if the real reason, or \textit{motivation}, for the anomalism of the mental is our \textit{unwillingness} to eliminate folk-psychology, the argument seems to be no good against those eliminative materialists who claim that elimination of folk-psychology is precisely what is wanted. It seems to me that Davidson would grant that psycho-physical laws are possible \textit{if} an elimination of folk-psychology can be tolerated as a consequence. In one of his earliest articles Davidson writes:

\[\ldots\] there is nothing that forces us to apply mental concepts.... It is partly an empirical question where and when we can meaningfully apply mental concepts, \textit{but it is partly a matter for our choice.} We can \textit{choose} to treat men under physical concepts, if we wish\ldots.\textsuperscript{1012}

If eliminative materialists think that they can explain all those aspects of human mental life which \textit{matter to us}, they can choose to describe humans in terms of physical concepts. But if this choice is made there should be no explanatory residues; I think we should insist that the eliminativist should explain those facts which we now explain in terms of mental concepts. The eliminativist conception should perhaps explain something more; why else would we be willing to adopt it?

It could be argued that the impossibility of psycho-physical laws follows from our view of what kind of being a human is. As Davidson writes: “The limit… placed on the social sciences is set not by nature, \textit{but by us when we decide to view men as rational agents

\textsuperscript{1011} Malcolm, 1971, 102.

\textsuperscript{1012} Davidson, 1964, 230, my emphasis.
with goals and purposes, and as subject to moral evaluation.”\footnote{Davidson, 1974b, 239. Davidson sees an interesting connection between “goodness” and “intentionality” because they are both supervenient on physical concepts without being definable in terms of them. He notes that this is not a coincidence because the concepts are closely related: “Moral evaluations are in the first instant applied to intentional actions”, as Davidson (1964, 47) notes.} The conviction that humans are free agents leads to the conclusion that the “nomological slack between the mental and the physical is essential as long as we conceive of man as rational animal.”\footnote{Davidson, 1970, 223.} Obviously, one cannot be forced to decide to view men as rational agents. According to Davidson, if we make this choice, we have to accept the vocabulary of propositional attitudes, because rationality depends on logical relations between the attitudes. He writes: “Thoughts, because they have propositional content, are unlike everything else in the world except for utterances in having logical relations to each other.”\footnote{Davidson, 2001f, 138.} Of course, this does not matter much to an eliminativist, who claims that human cognition, computation or memory have nothing essential to do with sentences or propositions or with the inferential relations between them. According to the eliminativist, a brain-centered conception which does not assign language an important role is now challenging the language-oriented conception of cognition.

It is clear that competing views about the nature of our self-conception exist, and what kind of conception one is willing to defend is, as Davidson claims, based on a decision. Paul Churchland notes that what is at stake in the debate is precisely our current self-conception which consists of “our shared portrait of ourselves as self-conscious creatures with beliefs, desires, emotions and the power of reason.”\footnote{Churchland, 1995, 18.} In his opinion, we should ask whether “our basic conception of human cognition [is] yet another myth…”\footnote{Ibid.} and whether “…a proper theory of brain function present[s] a significantly different or incompatible portrait of human nature?”\footnote{Ibid.}

The general motivation for the anomalism of the mental is thus clear. Commentators have, however, questioned whether the need to see humans as rational agents can be a reason to deny the possibility of psycho-physical laws. Antony notes: “If there are tight connections, there are tight connections – what can we do about it? If psychologists can begin to read our minds by scanning our brains, we will have to learn to live with it. It is not as if any of this is up to us.”\footnote{Antony, 1994, 236.} She recognizes, correctly in my opinion, that in the end it is up to us, according to Davidson. This explains why Davidson uses conditional expressions when dismissing the
possibility of psycho-physical laws. He claims, for example “to allow the possibility of [psychophysical laws] would amount to changing the subject.”

Antony argues that Davidson’s purpose is to block a certain possibility, namely, the competition between rational and non-rational evidence in the attributions of mental states. But as I have noted, Davidson is not denying the possibility of describing human behavior in physical terms. There is no real competition between neuroscientific evidence and folk-psychological evidence, because we first have to decide what kind of evidence we want to use, and this comes back to the question what kind of self-conception we want to hold. Eliminative materialists are entitled to their position if they are willing to accept its consequences. If one is willing to reject the view that mental phenomena should be described in folk-psychological terms, then the question of competition does not occur. It occurs if the Davidsonian criterion of mental is accepted and then the connections between physical and mental realities are being claimed. This being the case, the question of why our mental vocabulary is not an optional part of our conceptual resources (i.e. why this vocabulary is important and indispensable) becomes essential. The question of why a folk-psychological conception should be accepted is crucial when considering how to correctly understand the irreducibility of the mental. Here, so it could be argued, lies the ultimate and extremely important “competition” between a folk-psychological and neuroscientific conception of human beings. These understandings can surely compete, and therefore attention must be paid to the question why we want to preserve our everyday mental vocabulary and why it matters to us.

It could be claimed that the “anomalism of the mental” is far less principled than critics have assumed, because as Davidson notes: “The extent to which mental concepts fall short of being reducible to physical concepts measure the degree of anomaly.” The degree of anomaly depends on the degree of reducibility. The question about anomalism thus really comes back to the question of whether mental concepts or phenomena can be reduced to physical concepts or phenomena without bridge laws. The argument for the anomalism of the mental does not really work against the eliminativist, who is willing to reject the Davidsonian conception of the mental. However, I believe that the reasons for the irreducibility of the mental are such that an eliminativist cannot ignore them. It must be realized that Davidson

---

1020 Davidson, 1970, 216, my emphases.
1022 We should realize that the claim that our mental lives should be described in terms of folk-psychology is quite absurd, because we do describe our lives in these terms. However, in order to make clear the conflict between non-reductive materialism and eliminative materialism, a view that our self-conception could change is required. I think that currently we do not have reasons to believe that such a conceptual change could actually occur.
1023 Davidson, 1993d, 194.
would grant the possibility of the kind of bridge-laws whose existence an eliminativist or reductive materialist attempts to show. Churchland has claimed that the purpose of reductions is not to “try to establish, that there is any sort of necessary connection…” The modest purpose of a reductivist is not to claim that mental phenomena are “logically supervenient” upon physical phenomena or that the identity in question is a “metaphysical necessity” – or even that there is any form of “nomological” or “lawlike” connection between the two phenomena. Churchland points out:

[...] all of these diverse modal relations are philosophical extravagances or confusions imposed, post facto, on successful cases of historical intertheoretic reductions, all of which were achieved without the help of such modal relations, and none of which displays any of them.

So, according to Churchland the arguments of, for example, Chalmers, Kripke and Davidson are irrelevant for the question of whether mental–physical identities can be established. This kind of view seems to reflect more generally the view of those scientific philosophers who are looking for empirical solutions to philosophical problems. What this shows is that the reductions or identities that an eliminativist is looking do not meet the standards of certain philosophers. What this in turn shows is that the existence of these kinds of reductions or identities is something that a philosopher like Davidson is not denying, or at least does not have to deny. Davidson notes that certain philosophers may have in mind different standards of reduction than he does, and if this is the case, then the issue is mainly verbal. I think that certain misreadings may be a result of ignorance, and this is one reason why Davidson’s position must be carefully clarified.

Davidson is after for a much more fundamental and theoretical answer than Churchland when he asks whether there is some kind of important difference between psychophysical laws and other bridge-laws. Davidson is asking whether the mental–physical relation can be as tight as the connection between physical concepts. His question is whether there can be strict, exceptionless psycho-physical laws. Likewise, von Wright is asking whether mental–physical connections are necessary because he thinks that this question is what interests a philosopher. Von Wright’s conclusion is the same as Churchland’s: they both claim that the connections are not necessary but contingent. Malcolm thinks that this is a crucially important observation for the philosophy of mind, because it shows that materialism

1024 Churchland, 2005b, 191.
1025 Ibid.
1026 It should be noted that the “successful intertheoretic reductions” of which Churchland is talking are not cases of successful mental–physical reductions.
1027 Davidson, 1993d.
is a confused position. As he claims: “It is logically impossible that the claimed contingent identity of mental events with brain processes could be proved or disproved empirically. It is not a possible scientific hypothesis.” In Malcolm’s view the contingent relation cannot be philosophically interesting and it cannot be proven by science.

We can conclude that the original thesis of the anomalism of the mental does not rule out the kinds of identities that an eliminativist or a reductionist is searching for, and dismissing the possibility of these identities was not Davidson’s intention. Accordingly, a criticism which suggests that Davidson has not shown that there cannot be laws involving the mental and the physical simply misses its mark because, as Davidson makes clear in “Thinking Causes”, his purpose was only to establish that the mental–physical laws cannot be strict. He gladly admits the existence of useful *mental—physical generalizations*. However, whereas the *motivation* to argue against strict psycho-physical laws is clear and rejecting it casts doubt upon the argument for the anomalism of the mental, it is nevertheless interesting to consider whether sense can be made of the argument while rejecting the motivation for it. The argument for the anomalism of the mental resembles the argument of Bennett and Hacker that mental phenomena cannot be attributed to a brain, because brains do not and cannot fulfill the criteria in terms of which mental phenomena are attributed in the course of life. Bennett and Hacker claim that the use of concepts becomes nonsensical if the application conditions of concepts are violated. We have seen that according to Davidson, certain constraints govern the application of mental concepts, and dismissing the role played by these constraints leads to a change of subject and thence to conceptual confusion.

As we noted in chapter two, Davidson has a detailed view about the nature of the mental which is based on the conviction that the whole truth about the mental realm is captured in the process of interpretation. The constitutive roles of holism, externalism and normativity are essential for the argument for the anomalism of the mental. One could try to argue that these characteristics are *not* essential features of mental phenomena. For the moment, let us nevertheless proceed from the observation that Davidson thinks he has shown why and how holism, externalism and normativity *are* constitutive for propositional attitudes.

The important property of mental reality is that “the attribution of the mental phenomena must be responsible to the background of reasons, beliefs, and intentions of the individual.” We cannot attribute propositional attitudes to a person outside the framework

1028 Malcolm, 1971, 70.
1030 Davidson, 1970, 222, my emphasis.
of a theory of her beliefs, desires or intentions because “we make sense of particular beliefs
only as they cohere with other beliefs, with preferences, with intentions, hopes, fears,
expectations, and the rest.” Mental reality is holistically constructed and the relations in
this holistic web are influenced by the normative decisions of the interpreter. Malcolm can be
seen as defending a similar kind of view when he writes: “Our reactions, our natural attitudes,
towards the expressions, movements and utterances of people constitute a dimension of our
mental concepts. This dimension is contributed by the perceptive, responding subject.” A
situation where the constitutive principles of our mental vocabulary would break down and
we would still be using mental concepts is unintelligible. We mean and have to mean
something with our concepts, and to abandon the application conditions of concepts would
amount to changing the subject. If the constitutive principles changed, we would not be
talking about the same things as before, and we would not be using the same concepts
anymore. A reduction of the mental could lead to a change of our current mental concepts or
to changes in the ways we use these concepts. The question relevant for the argument for the
anomalism of the mental is that of the nature of our current concepts and their correct usage.

Because of their interconnected meanings, the constitutive principles of vocabularies
are a priori. Davidson claims that the conditions for the application of mental concepts which
he describes are necessary conditions. We hear: “Nomological statements bring together
predicates that we know a priori are made for each other – know, that is, independently of
knowing whether the evidence supports a connection between them.” If we know, a priori,
when the predicates are made for each other we presumably also know when they are not
made for each other. This, so it seems to me, also applies in the case of the mental and the
physical. Rationality and its governing normative principles are essential characteristics of the
mental, whereas the absence of rationality and normative principles is a characteristic of the
physical. If there were strict bridging laws between mental and physical concepts, the
characteristics of the mental which have no echo in any physical theory would be transmitted
to the physical realm, and vice versa. Because of the special features mental and physical
concepts, the connections between them have to remain.

Davidson claims that whether a statement can be considered to be law-like or not
depends on how the predicates in it can be paired or matched. We must consider not just

1031 Davidson, 1970, 221, my emphasis.
individual predicates, but the question of how different predicates can be related. As Davidson writes:

[... ] the fact that what inductions are reasonable concerning a class of objects... depends on the nature of the class. If this is so, it follows that we can’t simply say of a class or predicate or property that it is suitable for inductions; the reasonableness of the induction depends on the class (predicate, property) we team it with.\textsuperscript{1034}

The impossibility of strict psycho-physical laws is shown if we are able to demonstrate that mental predicates pick out classes of events which do not match up perfectly with physical predicates. This being the case, no precise laws between the predicates can be formulated.

If we accept that constitutive principles cannot be violated without creating nonsensical claims, it is easy to understand what Davidson means by saying that to allow the possibility of strict psycho-physical laws would amount to a change in the subject. With the existence of bridge-laws, attributions of mental states could be influenced by principles and evidence which are not part of the constitutive features of mental reality. What would be our reason for saying that it is mental states that we are talking about in this case? A concept would have been used outside its application conditions. Why would we think that it would still be the same concept? Wittgenstein once wrote about pain: “Pain has this position in our life; has these connexions; (That is to say: we only call ‘pain’ what has this position, these connexions)”\textsuperscript{1035} How could we cut of these connections and still claim that we are talking about pain?

Considering how difficult it would be to give up the idea that physical phenomena have a location in time and space may lead us to appreciate the strength of the constitutive principles of the physical. Imagine a situation where the axioms, laws, or postulates for the measurement of length would be rejected, but we would still continue to rank objects in terms of their lengths. What would be the meaning of the term “length” in this kind of situation? We cannot give up these axioms if we want to talk about lengths or rank objects in terms of this property. As Davidson notes: “The satisfaction of the conditions for measuring length or mass may be viewed as constitutive of the range of application of the sciences that employ these measures.... It is not easy to describe in convincing detail an experiment that would persuade us that the transitivity of the relation of heavier than had failed.”\textsuperscript{1036} He concludes: “If the law

\begin{flushright}
\textsuperscript{1034} Davidson, 1993i, 348.
\textsuperscript{1035} Wittgenstein, 1967, §533.
\textsuperscript{1036} Davidson, 1974b, 236-237.
\end{flushright}
of transitivity fails in a single case, the entire theory of measurement of length is false, and we are not justified talking of physical lengths.”

If we feel confident enough to say that there are constitutive principles of the physical which cannot be violated without changing the subject, why would we be reluctant to say this in the case of mental? What would be the reason to think that the constitutive principles of the mental would be “less constitutive”? According to Davidson, the constitutive principles of the mental seem to be stronger than the principles governing the physical. He says: “It is not merely, as with the measurement of length, that each case tests a theory and depends upon it, but that the content of a propositional attitude derives from its place in the pattern.” An isolated propositional attitude, which is not part of the rational pattern that an interpreter has to force on the interpretee, is nonsensical. Whether propositional attitudes capture the basic structure of human cognition is one question. But concepts in terms of which we understand each other have relations that cannot lose their relations to other concepts while remaining the same concepts. Without constitutive relations implicating other concepts, a concept is meaningless. We cannot make sense of certain concepts if they are detached from their ordinary usage.

An example of how the loss of such constitutive relations leads to an incomprehensible situation is well described by Stephen Stich’s case of Mrs. T. She is an elderly woman suffering from a memory loss that is getting worse and worse as the years go by. She remembers, for example, that president McKinley was assassinated; yet she cannot say whether McKinley is dead or alive. The question that Stich asks is:

Did she… believe that McKinley was assassinated? For just about everyone to whom I have posed this question, the overwhelmingly clear intuitive answer is no. One simply cannot believe that McKinley was assassinated if one has no idea what an assassination is, nor any grasp of the difference between life and death.

Stich’s general view is that:

[...] intuitive judgments about whether a subject’s belief can be characterized in a given way and intuitive judgments about whether a pair of subjects have the same belief are often sensitive not only to the potential causal interactions of the belief(s) in question but also to other beliefs that the subject(s) are assumed to have. The content we ascribe to a belief depends, more or less, holistically on the subject’s entire network of related beliefs.

1037 Davidson, 2001f, 131.
1038 Davidson, 1970, 221, my emphasis.
1039 Stich, 1983.
1040 Stich, 1983, 56.
Stich notes that this is the way that “intuitive judgments” work. When asked whether Mrs. T believes that McKinley was assassinated although she does not believe that McKinley is dead, most people will “intuitively” say no. But I would assert that concepts like “life”, “death” and “assassination” are just everyday concepts, and there is no more to their meaning than the intuitions of those who use these concepts. As Stich notes, the content of a propositional attitude depends on the subject’s entire network of propositional attitudes. This is what Davidson also claims when he emphasizes the constitutive role of holism for the identity of propositional attitudes. As he says:

[The] obvious logical relations amongst beliefs; amongst beliefs, desires and intentions; between beliefs and the world, make beliefs the beliefs they are; therefore they cannot in general lose these relations and remain the same beliefs. Such relations are constitutive of the propositional attitudes.1042

How could this view be challenged? Let us suppose someone is arguing that the content of a belief does not depend on its relations to other beliefs. What would the claim that “Mrs. T believes that McKinley was assassinated” mean according to this kind of view? If we try to answer this question, we have to refer to Mrs. T’s other beliefs. Most of us understand what it means to believe that someone was assassinated. If we hear that somebody believes both that McKinley was assassinated and that McKinley is still alive, our understanding fails. Situations like the case of Mrs. T where the connections between concepts are distorted show how obvious these relations are. We do not realize their importance in ordinary cases, but the consequences are dramatic when the connections are distorted.

Davidson stresses the importance of these conceptual connections in varying degrees. He notes, for example: “The contents of propositional attitudes are determined in logical part by their relations with the contents of other attitudes; to the extent that these relations of a particular attitude are broken or confused, the identity of that attitude is rendered less precise.”1043 Davidson writes also: “Each belief must be involved with other beliefs with which it is consistent if it is to be identified as having a clear content.”1044 Here the claim is that if the relations between attitudes are endangered, only the precision of the content is questioned. If a belief is not a part of a consistent web of beliefs, it cannot be identified as having a clear content. Mutual understanding may be endangered if the relations between an interprettee's attitudes differ from the relations between the interpreters’ attitudes. Confused relations do not, however, imply the non-existence of the underlying attitudes. This relaxed

1042 Davidson, 1985d, 351-352.
1043 Davidson, 1997b, 7.
1044 Davidson, 1997a, 125.
view about the nature of holism should be compared to the following: “Beliefs... are largely identified by their logical and other relations to each other; change the relations, and you change the identity of the thought.”

This is a stronger claim, suggesting that the identity of a belief changes if its relations to other beliefs change. The final conclusion is: “Many concepts are fairly directly connected, through causality, with the world, but they would not be the concepts they are without their connections with other concepts, and without any relations to other concepts, they would not be concepts.”

An isolated concept or propositional attitude would make no sense. Our concept of such attitude or concept would not make sense. It seems to me that sometimes Davidson goes too far when he tries to demonstrate how the holism of the mental should be understood. He claims, for example: “I can believe a cloud is passing before the sun, but only because I believe there is a sun, that clouds are made of water vapour, that water can exist in liquid or gaseous form; and so on, without end.”

Davidson gives a list of what one must believe in order to believe that a cloud is passing the sun. Similar suggestions are made when Davidson claims that in order to have the idea of what a snake is, you must believe things like “...a snake is an animal, it has no feet, it moves with sinuous movement, it is smaller than a mountain.”

To have the concept of a tree, you must believe that “they are growing things, that they need soil and water, that they have leaves or needles, that they burn.”

To have the concept of a spider, you must believe that “it is a living animal, that it is self-locomoting, that it has many legs, that it is apt to spin webs, that it must eat to continue living, that it will evade what it senses as dangerous....”

It is difficult to say how these kinds of claims should be interpreted. What Davidson should say, in my opinion, is that there is no fixed list of beliefs that one must have in order to have the belief B, and the lists given above are just examples of the kinds of beliefs that one must have in order to have a certain belief. But if this is the case, then it seems that we cannot give any examples of the specific beliefs that one must have in order to have a certain belief. It seems possible that one can have beliefs about clouds without having any beliefs about the properties of water, or that one can believe that something is a tree without knowing that trees burn. But this is certainly an issue where intuitions differ. Davidson’s examples show how strongly he refers to common-sense...
understanding as a determinant of the “correct” use of concepts. For example, the criterion of the concept of “spider” would perhaps be different for a biologist than what Davidson suggests.

I shall now clarify how the constitutive principles of the mental stand in the way of strict psycho-physical laws: Let us suppose that we have formulated laws of the form:

1) Necessarily, M occurs to person S at t iff N occurs to S at t.
2) Necessarily, M* occurs to person S at t iff N* occurs to S at t.

We also know that:

3) Necessarily, if M occurs at t, M* occurs at t.

From this, it follows:

4) Necessarily, if N occurs at t, N* occurs at t.

The example is borrowed from Kim.\(^{1051}\) In the example, M and M* are mental events which are connected to neural states N and N* by strict psychophysical laws (1) and (2). Statement (3) expresses a psychological entailment between M and M*, an entailment which is secured by our normative principles governing beliefs. Kim’s claim is that we could “read off” a physical law (4) from a purely psychological entailment (3), and this would mean that the transition from N to N* would be supported and secured by the normative principles of rationality and coherence. But this would mean that we would have transferred to the physical realm principles which have no echo in that domain, i.e. principles which have no application or relevance there?

Let us suppose that we could transfer the constitutive principles across schemes. We would attribute mental states according to the constitutive principles of the physical. This would certainly be a situation, warned about by Malcolm and Davidson, where the subject would have been changed. The attributions of mental states would not be done in terms of propositional attitudes but in terms of, for example, neuroscientific predicates and laws. It seems to me that in this kind of situation we would have no reason to say that we were talking

\(^{1051}\) Kim, 2003.
about *mental* states any longer, and since the mental aspect would have been dropped how could we talk about mental–physical connections either? In the connecting laws, phenomena mentioned on one side of the connective constrain the phenomena mentioned on the other side. This is one essential point of connecting laws, viz. to express dependencies between phenomena. If “x is M iff x is P” were true, then whenever there existed a physical state P, there would exist a mental state M. But, in the absence of additional mental evidence, what would or could convince us of this?

What would guarantee that P would always “underlie” M? An obvious answer is that the existence of a strict psycho-physical law would be such a guarantee. But what would be the reason to think that this law would hold for future cases? The law expressing identity between H\textsubscript{2}O and water is based on the observations of the structure of the substance which at a macrolevel is identified as water. But it is implausible to think that we could, in a similar way, observe the physical structure of mental phenomena? The question of whether mental phenomena can be found in the brain, whether they can be actually located in the brain is pressing when considering the question about the possibility of psycho-physical laws. An identity claim which is understood as a law must be based on evidence that explains the identity. Let us imagine that we doubted that each sample of water is a sample of H\textsubscript{2}O. To remove this doubt, we could literally check each sample and see that its structure is such that one molecule of water has two hydrogen atoms covalently bonded to a single oxygen atom. The identity claim “water is H\textsubscript{2}O” is just a generalization of the observations that samples of water have this molecular structure. What makes this identity claim non-problematic is that the molecular structure of water can be observed. The alleged mental–physical identity is problematic because mental phenomena are *certainly* not physical phenomena, as water without doubt is. Let us suppose that an identity claim that “M is P” has been established. Whereas the correctness of “water is H\textsubscript{2}O” can be tested in the above-mentioned way, how could we test the correctness of the claim that “M is P”? The claim itself would be an outcome of a complicated theory, the nature of which it is hard to even imagine.

The science which would provide such a theory would have to relate everyday psychological phenomena to states of the brain. But there is not even a *psychological* theory which would comprise the totality of facts about human mental life. Is an all-encompassing psychological theory possible? It would have to include all the mental phenomena we intuitively take to be mental, including all the possible beliefs that a person could hold. What kind of theory could this be? To make things worse, even if such a utopian psychological theory existed, there would still remain the tremendous task of identifying *all* the mental
phenomena in the realm of the brain. How this could be done in a way that would show that mental–physical identifications amount to identities and not to contingent token-correlations is beyond comprehension. How, for each and every mental phenomenon, a nomologically necessary and sufficient physical condition could be provided is a question that remains without an answer. In order to establish “bridge-laws” between the mental and physical domains, a comprehensive psycho-neural theory of the human mind would be required, but no one has even tried to suggest how such a theory could actually be constructed.\footnote{1052}

It may be tempting to think that the relation between a mental phenomenon and a physical phenomenon is the kind of identity that exists between water and $\text{H}_2\text{O}$; however, it is clear that the analogy cannot work. What enables us to claim that it is a law that “water is $\text{H}_2\text{O}$” is the fact that there exists a comprehensive theory, namely chemistry, which explains what this statement means and, as noted already, which enables us to check whether the claim is true or not. In order to show that “Donald believes that cats are animals iff his brain is in a state $P$” is true, a theory which would explain what it means for a belief to be a state of brain is required. The description of state “$P$” would have to be replaced with a description of a physical property of the brain. Would a description of the molecular structure of a certain brain state be the right kind of answer? The problem is that we simply do not have a view about what kind of description would be satisfying. Perhaps each level of explanation would satisfy some epistemological desires; but would it make sense to ask which level is the correct level of description? If you are interested in knowing “How Molecules Matter to Mental Computation”,\footnote{1053} then the molecular level could indeed be the right level of identification. But what if you are interested in knowing how atoms matter to mental computation? Then the right level of mental–physical identification would be the atomic and not molecular level. What would be the reason to stop here; would it not be reasonable to try to find the relevant correlates at the sub-atomic level?

It is not easy to see how mental states and brain states could have compatible identity conditions. Some, like von Wright and Malcolm, say that mental phenomena simply do not have spatial location. Brain states, on the other hand, have such a location. This means that states of the brain have a property that mental phenomena do not have, and vice versa. From this it follows that mental and physical phenomena do not have compatible identity conditions. This can be seen as a background assumption in Davidson’s argument against bridge-laws. Jeffrey Polger, who has recently defended an identity theory, claims that: “If we

\footnote{1052}{For this criticism see, for example, by Polger, 2004.}
\footnote{1053}{Title of Thagard, 2002.}
are certain that two things... have incompatible identity conditions... then they are not even candidates for identity."¹⁰⁵⁴ This resembles the view of Alan Sidelle, who discusses the question of how to distinguish the cases where there is a possibility of scientific identity from those where there is no such possibility.¹⁰⁵⁵ Sidelle points out, correctly I think, that in order to establish an identity between two things, we must already know something about the things in question. We must know whether the two things are similar enough, i.e. whether they share something in terms of which they can be seen as being identical. Polger has concluded that:

We do not know how to individuate brain states, properties, processes, events...Not only do we not how to individuate these things, we don't really even have a clue what such things are... as a matter of empirical fact at this time we do not know the identity conditions for brain states, properties, processes, events....¹⁰⁵⁶

One crucial problem is that philosophers keep talking about brain states without having a theoretical account of what a brain state is. But what exactly is a brain state? This question has not received much attention from philosophers.¹⁰⁵⁷ Can mental–physical identities be established if it is unclear with what the mental side should be identical?

On a Davidsonian account, mental events are physical events and physical events fall under strict laws; how can it be possible that mental events do not fall under such laws? After all, mental events just are the events which already fall under strict laws. An obvious answer is Davidson’s view that events fall under laws only when they are described in a certain way. Physical events, when described as mental, do not fall under strict laws. But what should be said of the following case? Let us suppose that we have formulated a law:

“Necessarily, if x is in M at time t he will be in P at time t₁”.

The law states that whenever a person is in a mental state M at time t he will be in a physical state P at time t₁. Since M is also a physical event, why couldn’t we formulate a law stating that “Necessarily, if x is in Pₜ at time t he will be in Pₜ₁ at time t₁” where Pₜ is a physical description of M? We can imagine that this could be possible. However, this possibility is of no interest with respect to the question of whether the fact that physical events are governed by strict laws implies that mental events fall under such laws as well. First of all, most obviously, Davidson’s claim is that there are no strict laws between the mental and the physical. “Necessarily, if x is in Pₜ at time t he will be in P at time t₁” is a law which mentions

¹⁰⁵⁶ Polger, 51-52.
¹⁰⁵⁷ For a discussion of this problem see, Brown, 2006.
only physical events, and as such it is a law the existence of which Davidson would allow. The existence of this kind of law is necessary if the PNCC is true. Second, the formulation of the law “Necessarily, if x is in P at time t he will be in P at time t_i” from the law “Necessarily, if x is in M at time t he will be in P at time t_i” obviously requires a way to actually reduce M to P. This, in turn would require a bridge-law of the form “x is M iff x is P”, which is not available. Thirdly, let us suppose that somehow we have formulated a law “Necessarily, if x is in P at time t he will be in P at time t_i” where P refers to a mental state. Well, since P is a physical state, it is clear that a change of subject would have occurred along the way, and we would have no reason to think the law as saying anything about a mental event at all.

The intuition on which Davidson’s argument is based is easy to follow. However, a critic could surely raise the following complaint. We have already agreed that every mental state is a physical state, a state of the brain or at least a physical state of the person. What kinds of connections exist between mental and physical states must be an empirical question. This being the case, it makes no sense to claim on a priori grounds that strict laws between the mental and the physical are impossible. Mental states are a part of nature and we cannot imagine what kinds of techniques for investigating nature might be possible in the future. We cannot say at the moment what kinds of laws the bridge laws between the mental and the physical will turn out to be. The fact that Davidson’s argument against the possibility of psycho-physical laws is a conceptual argument is surely a reason that philosophers with a certain kind of naturalistic attitude do not take it very seriously. A philosopher who does not want to take this kind of argument seriously thinks that we can draw a clear distinction between the way things are the way we talk about them. It is argued that we can find out how mental states are connected to physical states, or better yet, how they are identical with them, and whatever these answers turn out to be, they may be such that our current understanding of the nature of mental states has nothing to do with them.

To sum up how the disparate commitments of the mental and the physical prevent the formulation of laws between them, let us say the following. When we interpret the behavior of an individual, we attribute intentional actions which can be rationalized by citing the relevant beliefs and desires of the individual. This, the possibility of giving rational explanations, is the main purpose of our mental vocabulary. If beliefs and desires are to have this explanatory role they must, in virtue of their propositional content, provide the individual’s reason for engaging in the course of action in question, where that course of action is also propositionally characterized. In interpreting an individual we must attribute
those intentional actions and propositional attitudes that it makes sense, from our point of view, for the individual to perform or to have. This cannot be done without attributing at least a minimal degree of rational coherence to the interpretive. It is clear that the constitutive principles of the mental, namely, holism and normativity, are the key to the anomalism of the mental. The main characteristics of Davidson’s conception of the mental are also the main reasons for the anomalism of mental. The acceptance of the view that there cannot be strict psycho-physical laws requires that the Davidsonian view about the mental is also accepted; the anomalism of the mental cannot really be challenged without challenging the Davidsonian conception. The properties of mental states have no counterpart in our understanding of the world as a physical system. By the time of “Mental Events” Davidson’s views about the nature of holism and normativity were not fully developed; yet he clearly states them as the most important reasons for anomalousism. We are told:

What lies behind our inability to discover deterministic psychophysical laws is this. When we attribute a belief, desire, a goal, an intention or a meaning to an agent, we necessarily operate within a system of concepts in part determined by the structure of beliefs and desires of the agent himself. Short of changing the subject, we cannot escape this feature of the psychological; but this feature has no counterpart in the world of physics.1058

Or,

[...] the attribution of mental phenomena must be responsible to the background of reasons, beliefs and intentions of the individual...when we use the concepts of belief, desire, and the rest we must stand prepared...to adjust our theory in the light of considerations of overall cogency: the constitutive ideal of rationality partly controls each phase in the evolution of what must be an evolving theory.1059

Davidson concludes: “Clearly this holism of the mental realm is a clue both to the autonomy and to the anomalous character of the mental.”1060 What should be noted with interest is the remark that the nature of the mental realm is a reason why we cannot discover psychophysical laws. One could claim that this is not a reason against the existence of these laws and that Davidson is again confusing an epistemological and ontological question. This would be a fair challenge if Davidson accepted the clear separation of ontology and epistemology; ontology recapitulates philology. But we should also note that on occasions Davidson seems nevertheless to argue that there do not exist psychophysical laws. “We don’t know precise laws for explaining and predicting [psychological events]; but unlike the situation in the

1058 Davidson, 1974b, 230.
1060 Davidson, 1970, 217.
natural sciences, this isn’t because we haven’t discovered them yet; *it’s because there are no such laws.* ^1061^ This is an interesting claim because Davidson’s view is that “laws are linguistic” and therefore one should not be able to find them from reality. But if laws are not “out there” to be found, how can Davidson be so sure that there are no psycho-physical laws? What does it mean to say that in the natural sciences we can *discover laws?* Can these laws be discovered from “reality”? If the answer is affirmative, is this yet another difference between psycho-physical and other laws? Davidson has claimed that psycho-physical laws differ not only in degree but in kind from purely physical laws. ^1062^ His insistence that there are no psycho-physical laws bears a curious resemblance to Wittgenstein’s view from 1938:

> Determinism applies to the mind as truly as to physical things.’ This is obscure because when we think of causal laws in physical things we think of experiments. We have nothing like this in connexion with feelings and motivation. And yet psychologists want to say: ‘There must be some law’ although no law has been found… to me that there aren’t actually any such laws seems important. ^1063^

The psychologists of Wittgenstein’s times were insisting on the same ideas as the naturalists of our times are: there *must* be psychological or psycho-physical laws. Both Davidson and Wittgenstein deny this. There is also an interesting similarity in their views about why this is so. As we have seen, Davidson claims that certain features of the mental have no echo in the physical world; the interpreter brings to the situation a feature which has no counterpart in the world of physics. The normative element of interpretation forces us to consider what seems to be an intelligible interpretation from our point of view. But as Davidson notes: “Nothing in physics corresponds to the way in which this feature of the mental shapes its categories.” ^1064^ In his 1932–33 lectures, Wittgenstein considers how we could analyze the cause or reason of somebody’s laughter. He concludes that as far as person’s reason to laugh is concerned, “The success of the analysis is supposed to be shown by the person’s agreement. There is nothing corresponding to this in physics.” ^1065^ It is a peculiar coincidence that both Davidson and Wittgenstein use the same expression when they discuss the way how certain features of the mental do not exist in non-mental reality.

So are psycho-physical laws possible or not? I would strongly emphasize that Davidson is actually giving reasons *why we cannot allow* the possibility of these laws. We cannot allow the possibility of strict laws as long as we want to maintain a certain picture of

---

^1061^ Davidson, 1993e, 312, my emphasis.
^1062^ See Davidson, 1985 and Davidson, 1999n.
^1063^ Wittgenstein, 1966, 42.
^1064^ Davidson, 1995a, 6.
^1065^ Wittgenstein, 1979, 40.
humans and of mental reality. Whether we want to maintain it depends, among other reasons, on whether we take it to be useful. The possibility of psycho-physical laws entangles with the question about the prospects of “scientific psychology” and about the status of folk-psychology. The possibility or impossibility of strict psycho-physical laws is tied to the concept of human beings that we want to hold. It seems that certain kinds of psycho-physical laws are perfectly possible for a philosopher who would reject Davidson’s conception of the mental and, more provocatively, who would be willing to reject a certain image of man. Without a Davidsonian view about the mental — and without reasons to think that the preservation of our mental vocabulary is something valuable, that the motivation to defend anomalism is important — the argument for the anomalism of the mental does not really make sense.

Should we say that the anomalism of the mental is a true thesis? Davidson himself admitted that there was no proof that the thesis is true. I believe the most interesting part of the argument is the claim that we must reject the hope for strict psycho-physical laws as long as we take our mental vocabulary to have a function and purpose that cannot be captured by the physical vocabulary. Davidson notes: “[...] the application of mental concepts is governed by a different set of purposes, is anchored to a different weighting of evidence than the physical concepts.”1066 Whether these purposes are something that should be defended is a very complex question, and one which has concrete implications beyond philosophy. Do we want to defend a folk-psychological conception of ourselves or are we willing to hand over to experts the possibility of passing a final verdict about the mental nature of humans? The expert view would be based on physical, objective evidence. I think we must conclude that we cannot easily decide whether the anomalism of the mental is a correct description of the connections between mental and the physical. The best way to evaluate its plausibility is to see whether we would be willing to give up our mental vocabulary and what the consequences of the possible rejection of this mental vocabulary would be.

3.2 Why the mental has a priority over the physical

In the previous section, we reached the conclusion that there cannot be tight connections between the mental and the physical if each is to remain faithful to its respective evidential basis. This is the essential point of the anomalism of the mental. A critic would say that there

1066 Davidson, 1964, 230, my emphasis.
are no constitutive principles which could not be ignored, and would claim that the question about the nature of the mental–physical relation is empirical. But we also saw that there is a strong pragmatic element in the argument for the anomalism of the mental. Davidson claims that we cannot allow the possibility of strict psycho-physical laws; the limit placed on the social sciences is a limit placed by us. I think that with these claims Davidson takes a stance on the question of how we should react in a case where rational and non-rational evidence would compete around the question what kind of states can be attributed to a subject. To claim that non-rational evidence cannot generally override rational evidence is to claim that ultimately mental evidence has some kind of priority over physical evidence. This in turn suggests that mental reality enjoys a special kind of autonomy over physical facts.

Von Wright strongly stressed this point when he defended the autonomy of psychology and the need for a psychological understanding of a human being. I believe that von Wright’s views about the priorities between vocabularies provide a very useful picture, which helps to clarify the sense in which mental phenomena are autonomous in relation to their physical counterparts. My claim is that the von Wrightian view clarifies also Davidson’s views about the anomalism as well as the Davidsonian claim that any attribution of mental phenomena must remain faithful to its proper source of evidence. Von Wright’s view clarifies how the status of constitutive principles can be understood. I claim that he makes an important contribution to the current debate about the status of non-reductive physicalism.

We can consider the strength of the criticism that there are no constitutive principles constraining mental reality and the subsequent claim that the question about the nature of mental–physical laws is an empirical question by considering the following thought experiment, which gives a concrete example of the autonomy of the mental. Through it we can appreciate why the mental has a certain priority over the physical. Let us suppose that at birth a child were to be transferred to a laboratory environment. He would grow up to be a somewhat normal, thinking adult. With techniques currently only imaginable, the development of the child’s brain would have been scanned and “recorded” so that we would have an extensive “map” connecting this person’s propositional attitudes and other conscious mental phenomena with the states of the brain. We can imagine that with this kind of perfect map, scientists would be willing to say that whenever the child thinks that x, the physical state P occurs, whenever he wants y, a physical state P_a occurs, whenever he is in pain, a physical state P_b occurs, and so on. The constructing of this kind of map would obviously be a colossal project and it may be the case that such a project is not humanly possible simply because of
some empirical restrictions. Let us nevertheless imagine that with the available future super-technology, the construction of the mental–physical map would have been achieved.

Let us now suppose that the person, Georg, whose brain states and mental phenomena have been correlated, is being observed by two groups of scientists. The first group watches Georg as he *lives* in the laboratory environment. The second group consults their computers, which provide detailed information about the *brain* of Georg and interpret this information according to the existing map. Group one predicts and explains the behavior of Georg based on his linguistic and non-linguistic behavior. In doing this they rely on their general knowledge of the human nature, which is the kind of knowledge that each of us already has. The second group makes their predictions and explanations based on what they know about the relations between brain states and mental phenomena. The question that I want to raise is: what is second group’s reason to attribute a propositional attitude to Georg? It is the fact that a computer indicates that a brain state P occurs in Georg at time t. Can the second group be more certain in its opinion that a specific propositional attitude is present than the first group? Without seeing and interpreting the person, would they be confident enough to say, for example, that Georg is intentionally trying to do X, say, to scratch his back? Would they be able to distinguish an intentional scratching from a non-intentional one? If yes, it would have to be the case that intention is something that can be found from the brain. Would they be willing to identify actions by reference to brain states only? Would they be willing to say that if two people are identical, brain-wise, their mental lives *must* also be? These are questions without answers because currently we can only imagine or speculate about what the answers could be. But the questions show the extent of the problem. What can be said is that if the verdict of the second group were to override the verdict of the first group, if we were to trust the computer over our own skills of interpretation, then this would indicate a major change in the way we think about the nature of human life. Perhaps such a change is possible, but it is certainly not possible without drastic changes in all areas of life.

Both von Wright and Davidson considered the possibility that we would know the exact physical details that happen in a person when a mental state occurs. They considered the kind of possibility being realized that the example of the laboratory and the two groups of scientists is meant to describe. As I noted in section 2.5.4, Davidson’s conclusion was that even if all the physical details of the situation were known, we could not avoid the need to interpret the behavior of the person if we were willing to know what the person meant or believed. It is interesting that at one point von Wright seemed to think otherwise. He claimed that it is “logically possible” or “conceivable” that a complete story of an agent’s action could
be “read off” from a full description of its counterpart in the person’s nervous system. In the end von Wright rejected this possibility and ended up defending the autonomy of the mental level.

Perhaps our intuitive view is that the majority of human actions could not be explained or predicted by studying the brain in isolation. Whereas the neural cause of a bodily movement could in principle be found out, in most cases the question of what a certain bodily movement means would still remain open. Actions are context-dependent; it is not the case that actions are merely identified in relation to the actual life in which they are embedded, the situational factors are constitutive for certain actions. For example, Georg cannot sign a death-sentence if his position in the society does not give him an authority to do so. Georg can be thinking that he is signing a death-sentence, but if at the time of signing, unbeknown to him, he is deprived of the authority to do so, he is not signing a death-sentence. Whether or not somebody is intentionally doing something in a given case cannot be answered straightforwardly without considering the act in the context in which it takes place. Even if intentions to do something could be located in the brain, the question of what kind of actions these intentions actually produce cannot be answered by studying the brain. The bodily movement that is the result of a neural cause counts as a complex action only if the appropriate cultural, social and legal institutions exist in terms of which the action can be understood. As von Wright has noted a very simple movement like Georg’s raising his hand is a poor example of human action because most things that people do are such that they cannot be described by describing physiologically or neurophysiologically what goes on in the person’s body at the time of the action. If O-physicalism is true, it is plausible that there is a physiological and neurophysiological description of what happens when Georg’s hand is rising, or even what (physically) happens when Georg is raising his hand. But what the raising of the hand means depends on factors which must be located outside the brain of Georg.

It is of some interest to note that scenarios where an outsider can predict the thoughts and consequent behavior of others have become increasingly popular in recent works of

---

1067 See von Wright, UPg. This is a curious claim because at times von Wright claimed precisely that the meaning-constituting relation between behavior and mental phenomena is conceptual.

1068 It could be claimed that what matters is only whether Georg thinks that he is signing a death-sentence, because there is no difference in Georg in the cases when he is genuinely signing such a sentence and when he is not. We would hold Georg responsible of his signing even if it were the case that, unbeknown to him, his authority was deprived at the time of signing. Whereas it is true that there is no difference in Georg in the cases when he is really signing a death-sentence and when he is not, this merely shows that if we want to understand actual human behavior and its consequences, physiology alone cannot help.

1069 Putnam’s familiar twin-earth scenarios, Davidson’s swampman, and other similar thought experiments are meant to show that mental states must to be identified by reference to factors outside the agent. The “externalism” of mental states is an essential feature of them.
popular-culture. In the imagined scenarios, the possibility that an outsider could know these facts before the subject knows them is often entertained. These scenarios are familiar from science-fiction movies, but popular culture often reflects the intellectual climate of its time. One can speculate that the reason why these kinds of imaginings are currently so popular is that scientific evidence is interpreted as pointing to the direction according to which mind-reading may be possible some day. The thought-experiment involving the two groups of scientists is therefore perhaps not as absurd as it may sound. In fact, if we go through the recent issues of *Newsweek* the picture described there is quite contrary to the view about the mind described for example by Wittgenstein, Malcolm, von Wright, and Davidson. A brief analysis of the current literature shows that neuroscientists are already considering the practical problems of mind-reading, like the question of whether the subject’s compliance is required in order to draw conclusions from the study of the subject’s brain.

Scientists interviewed in a *Newsweek* article titled “Mind Reading Is Now Possible” claim: “The new realization is that every thought is associated with a pattern of brain activity and you can train a computer to recognize the pattern associated with a particular thought”\(^{1070}\); or we hear: “If our approach could be expanded upon, it might be possible to predict what someone was thinking or seeing from brain activity alone.”\(^{1071}\) The author of the article, who I assume has not considered Davidson’s argument against strict psycho-physical laws, concludes: “If what your brain does when it thinks about an igloo is almost identical to what mine does, that suggests the possibility of a universal mind-reading dictionary, in which brain-activity pattern \(x\) means thought \(y\) in most people.”\(^{1072}\) It is interesting that in an article which is meant for a wide audience, many of the mistakes or confusions which we have discussed in the previous chapters are being made. The author uses the expression “almost identical” and draws conclusions about mental–physical connections from it. But what kind of identity is this “almost identity”? It is being claimed, *pace* the arguments of Bennett and Hacker, that the brain “does” something and that it “thinks”. It is also claimed that brain-activity pattern \(x\) *means* thought \(y\) in most people. This suggests that a similarity of brain-states is used as a *criterion* for the sameness of thoughts. This means that the relation of meaning which now exists between mental phenomena and behavior has changed to something else. These are examples of how the conceptual confusions of modern neuroscience influence writings which reach a wider audience. It is these more popular views

\(^{1070}\) Neuroscientist John Dylan Haynes quoted in Begley, 2008 (my emphasis).

\(^{1071}\) Neuroscientist Geraint Rees quoted in Begley, 2008.

\(^{1072}\) Begley, 2008.
which will have an impact on the way that people start to think about the mind–brain relationship.

These kinds of confused suggestions can have an effect on society, and they should interest a philosopher of mind who does not want to alienate himself from the world. Von Wright’s opinion was that the best thing a philosopher can do is to act as a critic of his own time. Here the task of a philosopher is to point out the obscurities in the ways in which the mind–brain relation is currently expressed. The Newsweek reporter cannot be blamed for drawing these kinds of philosophically problematic conclusions, given that neuroscientists have recently published articles like “Decoding Mental States from Brain Activity in Humans”\textsuperscript{1073}, “Reading Hidden Intentions in the Human Brain”\textsuperscript{1074} or “Unconscious Determinants of Free Decision in the Human Brain”.\textsuperscript{1075} These scientific articles are followed by more popular articles explaining the results of the research for the public. The titles are of these popular articles are telling: “Mind Reading Machine Knows Your Thoughts Before You Do”, “Brain Scanner Predicts Your Future Moves”, “Mind Reading Machine Knows What You See”. These are not the topics of science-fiction movies, but the topics of articles from the publication New Scientist from the years 2006–2009.

If we listen to the conclusions of neuroscientific research, there are serious reasons to doubt that, for example, Davidson’s claims about token-irreducibility, Malcolm’s claims about the nature of dreams and dreaming, or von Wright’s claim about the nature of intentions are true.\textsuperscript{1076} From the perspective of modern neuroscience the claims of philosophers are not very convincing. According to Haynes it seems that every thought is associated with a certain pattern of brain activity, and a computer can be programmed to recognize this pattern. This casts doubt upon Davidson’s claim that beliefs are not little entities lodged in the brain. It is being claimed that functional magnetic resonance imaging sheds light on how the brain processes visual information, and how “it might one day be used to reconstruct dreams”.\textsuperscript{1077} This goes blatantly against Malcolm’s views on the nature of dreams. Von Wright claimed that intentions cannot be found from the brain and that one cannot perform tasks by sheer will; yet the existence of so called brain-computer interfaces already show that these claims are controversial. As I have noted in chapter two, there is an enormous empirical pressure against conceptual arguments in the philosophy of mind if and when these arguments go, as

\textsuperscript{1073} Haynes and Rees, 2006.
\textsuperscript{1074} Haynes et al, 2007.
\textsuperscript{1075} Soon et al. 2008.
\textsuperscript{1076} Malcolm, 1959.
\textsuperscript{1077} Singer, 2008.
they sometimes do in the case of Davidson and von Wright, beyond clarifying what makes sense. When competition between an empirical claim and a philosophical argument occurs, empirical evidence is relevant. That is why, for example, Davidson’s and Rorty’s arguments against token-reducibility are so problematic.

But are neuroscientists taking conceptual short-cuts in their reasoning? It is telling that in the recent neuroscientific articles the discussion focuses on hidden intentions, unconscious determinants or brain activity from which mental states must be decoded. If philosophers have ignored empirical results, it seems that neuroscientists have, likewise, ignored the conceptual points made by philosophers. Are the concepts of a “hidden intention” or an “unconscious determinant of a free decision” coherent? Concrete examples of some of the problems which may trouble a philosopher but which are irrelevant for a neuroscientist occurred to me when I was a volunteer in an experiment where the functioning of a brain-computer interface was tested. A brain-computer interface (BCI), sometimes called a direct neural interface or a brain-machine interface (BMI), accepts commands directly from the human or animal brain without requiring any physical movement of the subject, and it can be used to operate a computer or other technological equipment. What a few years ago would have sounded like science fiction, namely controlling computers with your brain, is now a real possibility. Von Wright, for example, seemed to doubt whether this could be possible, although in his later writings he had to admit that current empirical results “look interesting”. As he says, commenting on the mirror-neuron research on monkeys: “these findings about the monkeys and their reactions to what (a man does), they are very, very interesting and it’s an advance towards not only speculative ideas about correlations, but also findings about generalities.”

It seems to me that it is appropriate to question whether this admittance puts the theory of action and its autonomous rational explanations into a very challenging dialogue with neuroscience. Von Wright’s admission that empirical research may produce generalities puts his view also in an interesting relation to type–type theory.

As a result of discovering mirror neurons, neuroscientists can literally say what a monkey is doing just by looking into its brain. The kind of “map” imagined in our laboratory-

---

1078 For information about the research group and their publications see http://www.lce.hut.fi/research/css/bci/.
1079 For results, see Kauhanen, Nykopp and Sams, 2006.
1079 Von Wright (UPg) concluded that man cannot “make things happen” by “sheer willing”. But the way BCI works comes, in my opinion, very close to a situation where a person makes things happen by sheer willing. Von Wright claimed that if somebody could make things happen by sheer willing, the result would not be anything that we would call action. As far as our current concepts are concerned this is perhaps true; but can this fact remove the possibility of producing effects by sheer willing?
1080 As quoted in Petit, 1999, 121-122.
1081 This challenge for von Wright’s position is raised by Petit, 1999.
example seems to be, in a modest form, a reality in the case of rats. In the human-case, BCI
can be used to connect brains to computers. Current interest in this kind of research and
technology is understandably high. The possibilities of BCI-applications are, at least in
theory, numerous. People who cannot speak or move could communicate through computer.
Humans could control machines, like robots, directly with brain activity. People who are in a
seemingly unconscious state could perhaps be reached through a direct link to their brains. All
these suggestions and many others have been put forward in the scientific literature in the past
few years. It seems that since the technology which enables you to move a cursor on a
computer screen with your brain activity has already been invented, the possibility of
performing more complex tasks is, more or less, just a matter of engineering. Hayden and
others conclude that their work on the neural correlates of intentions has:

[...] important implications not only for the neural models of executive control, but also for technical
and clinical applications, such as the further development of brain-computer interfaces, that might now
be able to decode intentions that go beyond simple movements and extend to high-level cognitive
processes. This is perhaps not surprising. If the mind can be understood as being a physical “thing” and if
it has a causal role in controlling behavior, it is no surprise that the link between the mind and
the outer world could be replaced with a BCI. We can already control the cursor on the
computer screen “with our mind”, the concrete physical link between the computer and our
mind being our hands! Epiphenomenalist worries aside for the moment, our minds can
already affect the world; BCI only makes the link more direct. We could therefore defend von
Wright’s skeptical claim that a person does not make things happen by sheer willing; she
produces brain states which a machine interprets and makes certain things happen as a result.
Let us imagine that the machine could be completely left out of the picture. A person could
make something happen by sheer willing. Would we claim that in this case too there must be
some mediating link between the willing and the consequence? If we insist that such a link
must be found, then it indeed seems to be the case that sheer willing cannot have
consequences. But why demand that there must be a link?

Actual participation in neuroscientific research was illuminating because it made me
realize that scientists were not troubled at all by the question of whether the phenomenon
which they were measuring was a “state” or an “entity” and they were not very interested in

---

1082 Tamburrini, 2009.
1083 Hayden et al., 2007, 326.
the question of what the exact relation between the mental and physical aspects of my concentration was. They were not interested in the question of whether the mental aspect was identical with the physical aspect in the philosopher’s sense of identity. Polger, who is willing to defend type–type identity theory on philosophical grounds and has also participated in some research in which his brain was scanned, draws the conclusion that “Each picture or kind of picture is at best an incomplete representation of what is going on in the brain. None of them gives pictures of ‘brain states’ that could play a general explanatory role in science of the brain.” 1084 Given the recent claims concerning what can be found from the brain, it could be tempting to claim that the scientists studying their computers could indeed predict and explain everything that the group studying Georg in an old-fashioned way could and, moreover, that they could do this with absolute accuracy or at least with much higher accuracy than the group without computers. If we assume that physical events are “deterministic”, it is tempting to maintain the vision that prediction in terms of brain states, which are physical states, could be much more accurate than prediction in terms of familiar mental concepts which refer to physical states of queer nature. Before we accept this vision, I think we should consider what von Wright says about this possibility. He suggests an interesting way in which mental reality could enjoy autonomy over the physical.

Let us reconsider the laboratory example. Neuroscientists already claim that mind-reading machines can know your thoughts before you do, and that a brain-scanner can predict your future moves. In our laboratory example, the “map” which would be used by the scientists studying Georg through computers could be something like the “universal mind-reading dictionary” the future existence of which was predicted in Newsweek. To be really useful, the map should contain laws and not just descriptions stating regularities between two kinds of phenomena. As Davidson claims: “[…] what we want as a law of correspondence is not an accidental correspondence, but one which, we have reason to believe, will apply to cases which we have not yet examined.” 1085 Davidson admits the possibility of non-strict laws of correspondence, but their existence is hardly a reason to think that new instances of the “same” mental phenomena would have the physical counterparts predicted by the non-strict law. A map without strict laws would provide knowledge about mind–physical relations, but how could we trust the predictions made according to it? How could we trust that these regularities are firmer than the folk-psychological regularities that are already known? As Kim has observed, the claim that A is an attribution condition for p is more than the

1085 Davidson, 1964, 47.
affirmation of a mere de facto coincidence of A with p; to assert it is to commit oneself to a statement with modal force. Let us suppose therefore that the map tells, for example, that “Necessarily, if C obtains P occurs” as well as “Necessarily, M occurs iff P”. Here “P” is a physical state, “C” describes the conditions under which we can say that P occurs and “M” is a mental state. It is easy to make sense of the idea that the occurrence of condition C is decisive for P. With the existence of a law “Necessarily, M occurs iff P”, we should draw the conclusion that the occurrence of certain set of physical conditions C settles, beyond doubt, the question of whether M occurs. In other words, C has become a determining condition for M and C is thus a criterion for the presence of M. This is what happens in the case suggested by Newsweek, where brain-activity pattern x means thought y in most people.

Kim’s interpretation of Davidson’s argument against strict psycho-physical laws concludes that with the existence of such laws the criterion of the mental would be changed to something which, according to Davidson, cannot be the right kind of criterion. Neural attribution conditions cannot be the right kind of attribution conditions for mental phenomena. This conclusion can be drawn if we consider what the actual attribution conditions for mental phenomena are and how mental phenomena are attributed in the course of real life, outside laboratories. The importance of actual attribution conditions have been stressed by Malcolm, von Wright, Bennett and Hacker who, following Wittgenstein, make much use of the distinction between a phenomenon being a symptom of a property and a phenomenon being a criterion for the presence of a property. Although Davidson does not use these Wittgensteinian terms, I believe it is easy to see how his argument relates to the one described below. When we consider the nature of the mental one crucial question is: How does a belief manifest itself? How does it show that a person has a belief? Von Wright’s answer is: “In that I am likely to answer questions ‘Yes’ and ‘No’, to do certain things and refrain from doing others, make some preparations, warn people, etc. ‘This is how my belief shows itself’.”

Starting from the third-person perspective: “We attribute to a person a belief on the basis of complex pattern of his bodily reactions (and ‘dispositions’ to react). This is a conceptualization of observations under an aspect of intentionality.” Here we see both the ways in which a belief shows itself and how beliefs are attributed to others. As I claimed in chapter two, the connections of these claims to Davidson’s views about the nature of belief and belief-attribution are obvious. According to this view, to “have a belief” is more like something we do than something we experience and the having of a belief is nothing physical.
inside the brain in any meaningful sense. But is behavior merely evidence for the existence of belief, is it a symptom of belief? If so, what precludes the possibility that there could be other kinds of evidence for mental phenomena such as descriptions of the neurological states in which the subject without doubt is? The idea of mind-reading machines and brain-scanners is clearly based on the view that neurological states would be good and in many cases conclusive evidence for the attributions of mental phenomena.

Von Wright, however, claims that behavior is not merely evidence for the existence of a belief. To say that it is, would be to say that behavior is a symptom of having a belief. But, according to von Wright, this is not the right way to understand the relation between mental phenomena and behavior. He describes the difference between symptoms and criteria in the following way:

The connection between a symptom and that of which it is a symptom is “empirical”. This means, roughly, the following: if A is a symptom of B, the existence or occurrence of A may make us anticipate, expect or predict, the occurrence or existence of B. But whether B actually occurs or exists will have to be established on independent grounds, i.e. on grounds which do not themselves make appeal to (the occurrence or existence) of A. These independent grounds, moreover, are sometimes, but not necessarily, what we call criteria (as opposed to symptoms) of B.

The connection between a symptom and that of which it is a symptom is empirical and can thus be established by research and observation. It can be established only in this way. The connection between a criterion and that of which it is a criterion is instead conceptual, logical or semantic. In this context, these three terms are roughly synonymous and if there is a difference between them von Wright does not discuss it. The connection between a criterion and that of which it is a criterion is not empirical; it depends of what we mean by certain concepts. Once the meaning of concepts is “settled” in a given linguistic community, the conceptual, logical or semantic relations between concepts are established. This is a crucial point, the importance of which cannot be exaggerated. It is, I think, impossible to prove that this is so; it is something which has to be accepted. By this I mean that it is a starting point which cannot be further grounded. Yet, this starting point is not difficult to accept if we consider the way that linguistic practices actually work. Here it is important to recall the Wittgensteinian suggestion that philosophical problems occur when language goes on holiday. The actual usage of language is a standard against which the meaningfulness of

---

1088 I challenged these views in chapter two. Here I use them to establish the point of von Wright and Davidson, but I am not accepting them as true.
1089 Von Wright, 1998, 110.
1089 The difference in their basic points of view is what makes the dispute between Bennett / Hacker and Dennett / Searle so severe. The basic disagreement is well described by Churchland (2005a, 464) when he notes that the participants have “a different set of convictions and a very different intellectual agenda….”
statements must be compared. Without any standard, the possibility of agreement, disagreement and discussion is lost. We do not have to accept that this has to be so; it is merely one attitude towards the nature of philosophical problems. If this attitude is accepted we should never forget Wittgenstein’s question: “[…] is the word ever actually used in this way in the language-game which is its original home?” 1091 This question should be asked frequently, especially among philosophers who start to create conceptual structures which lose touch with reality. Anything goes if there are no standards. Various things can be “proved” by readjusting ordinary language.

As I have shown, Davidson’s claim is that it is possible to know a priori which predicates are made for each other. I argue that the reason for this is precisely the fact that the criteria for the application of predicates are conceptual, logical or semantic. In this respect, the question of whether strict laws between the mental and the physical are possible is wholly a conceptual and not an empirical question. In the case of belief and behavior the presence of behavioral criteria does not only indicate but in fact means that the subject in question believes something; the behavioral criteria of B determine the meaning of the term “B”. As von Wright says: “the presence of the behavioural criteria of expectation or of curiosity means that the being expects or is curious about something”. 1092 Moreover: “The behavioral reactions to sound are constitutive of hearing. Their presence (occurrence) means that the subject hears something”. 1093 Von Wright claims that behavior has a constitutive role in the attributions of mental phenomena. If behavior has this role with respect to mental phenomena, if it is a criterion, then the question of whether mental phenomena can be attributed in the absence of behavior simply does not make sense. The question is wrongly stated and it becomes an empty question. If there did not exist behavior which is logically or conceptually characteristic of mental states, the statement that a subject has a mental state would not make sense. These statements could make sense if the concepts of mental phenomenon or behavior were something completely different than they actually are. As things stand with our current concepts, the relation between behavior and mental phenomena is essential.

Von Wright is making a point about the language of the mental. On the one hand, there exists the language game of the physical, and on the other hand, the language game of the mental. The concepts used in these “games” show, so to speak, what kind of functions they perform. When we say, in the language game of the mental, that Georg is curious we are

1092 Von Wright, 1998, 146.
1093 Von Wright, 1998, 161, first emphasis mine.
not, in the first place, indicating that such and such neurological processes are occurring. That indication would be an irritating change of subject. That we are not indicating this can be verified by consulting the actual users of such language. In our linguistic community, “being curious” simply is not a property of the brain. When attributing curiosity to a person, we do not mean that her brain is in a certain state; we mean something very different. Although “being curious” is a state or process the exact mental nature of which is unclear, we can use this expression in a meaningful way and without problems most of the time.

A neurological process is not what we mean by the concept “curious” and the attribution of curiosity does not depend on anything that we know about the subject’s brain. The language game of the mental is structured around different kinds of phenomena than what take place in the brain. It is not possible to specify the exact conditions which must prevail for us to attribute a specific mental phenomenon to a subject. Mental phenomena are part of a broader context, and they manifest themselves in a certain kind of “form of life” that is familiar to us. This form of life is characteristic to humans, and we are familiar with it because we grow up in a society in which this form of life is naturally expressed. Von Wright notices that: “In attributing reasons for action to an agent we normally also attribute to him various abilities, beliefs, desires and inclinations, the understanding of institutions and practices of the community, and other things which characterize him as a person”. This claim is a remark about the application conditions of mental concepts. The attribution of mental concepts is impossible if the application conditions referring to complex social factors are lacking; attributions make sense only if the institutions and practices of a community exist, as von Wright claims. These are, for example, the institutions of language, morality and law as Davidson notes. Our mental vocabulary is so deeply embedded in our actual life-practices that the concepts would become meaningless if they were detached from their everyday context. Likewise, without these concepts the majority of our life-practices would collapse. Institutions of law and morality, as we currently understand them, would become meaningless if we were to stop describing human behavior without reference to mental phenomena like beliefs.

If interpretationism is taken seriously, we should conclude that an attribution of a mental state to a subject is not a description of the inner state in which she is. The attribution is rather meant to be a description of a person as a participant in a certain kind of activity, or

---

1094 Von Wright, 1998, 27.
1095 This is one reason why eliminative materialism is so counter-intuitive a position. As Stich (1993) has noted, if we are forced to give up our mental vocabulary we have to give up disciplines like economics, political science, sociology, anthropology and all the other areas of inquiry that depend essentially on mental concepts.
in a form of life. To say that the subject has mental states is to imply that she is a creature who can participate to certain kinds of activities; she is a creature who can live a certain kind of life. The study of the logic of the language of the mental amounts to noticing, and then describing, how, when and in what circumstances we actually make, can make, or *are allowed to make* psychological ascriptions to others or give expressions of our own psychological situation. It could be claimed, as Davidson, von Wright and Wittgenstein do, that to describe the logic of our mental language is also to describe the “metaphysical conditions” of thought. To insist that in describing these situations and in studying the logic of the mental language we are not describing the “real nature of mental phenomena” is to assume that there is an independent nature of the *mental* which could be separated from the ways in which mental language is actually used and which could be studied outside this language. What is the rationale for believing this? If participation in a certain form of life is a *criterion* that a creature must satisfy in order to be credited with mental phenomena, then a suggestion that mental phenomena could be studied outside this context does not make sense.

How does the distinction between symptoms and criteria relate to the question of the possibility or impossibility of strict psycho-physical laws? Let us start from the common-sense observation that we have knowledge of mental states which does not depend on our knowledge of the brain. I think it is indisputable that basically all our common-sense knowledge about mental states is independent of our knowledge about the brain. Humans used mental concepts before they knew anything about the properties of the brain, in fact, before they even knew that they had a brain. This important observation shows that the application conditions, the *criteria*, of mental concepts must – in our system of concepts – be something which has nothing to do with the fact that humans happen to have a brain, which happens to be the organ where “thinking takes place”. Neural discoveries about the mental, if it even makes sense to speak in these terms, are irrelevant for the application of mental concepts. As von Wright says: “We have these [mental concepts] and know how to handle them in linguistic communication…. Unless we had the concepts we should not know how to identify the phenomena falling under them….” One can compare this to Davidson’s remark according to which “if we never understood anyone, the concepts of language, understanding and thought would have no application for us.”

---

1096 According to the conclusions of section 2.5.4 we should perhaps say that the organ in which thinking takes place is nonsense. However, we can at least say that, most likely, people could not think if they did not have a brain. Brain is required for thinking but from this it does not follow that thinking can be located to the brain.

1097 Von Wright, 1998, 93.

1098 Davidson, 1994, 127.
mental concepts, that we know how to apply them, and that we had them before any discoveries about the brain were made is a reason to say that these concepts, and the phenomena they designate, have their own non-neural criteria. Because the criteria of X determine the meaning of the term “X”, it is not possible that an empirical discovery could change the meaning of X. If an empirical discovery would change the meaning of X, then it would not make sense to say that we would be talking about X anymore; we should say that the discovery would not change the meaning of X it but it would introduce a new concept X₁. Von Wright notes, for example, that “[…] a person cannot correctly be said to have pain, if he does not feel it and does not exhibit pain behaviour.”¹⁰⁹⁹ He goes on:

…a “clear case of pain”; a case in which I necessarily know, “by introspection”, whether I have pain or not. Such cases… are the normal or typical cases; if they were not this, then our concept of pain would be different from what it is now…. Because my certainty, in a “clear case”, that the sensation I have is one of pain is my certainty that I know (master) the correct use of the word for it (in English “pain”).¹¹⁰⁰

Von Wright is describing our current concept of pain. We can suppose that someone, like an eliminative materialist, would deny this view about the nature of the concept of pain and thus about the nature of pain. Antony, far from categorically denying this view, nevertheless notes that “at some point, we may well allow that, strange as it may seem, people are sometimes wrong about whether or not they’re in pain.”¹¹⁰¹ It is unclear what the statement “He suffers pain” could mean to a critic who rejected von Wright’s view about the nature of pain. It is also unclear how successfully he could use the concept “pain” in discussions with others. He could not use it in a way that would fit with our intuitions about the nature of a pain. This being the case, there would be no reason to say that an eliminativist would be talking about the same thing as we are. We would have to ask of a person who denies von Wright’s view about pain what his criteria for pain are, and why we should think that those criteria are criteria of pain. Since we would still understand pain according to the old familiar criteria, what would be the reason to think that his criterion of pain would apply to the same phenomena as ours? Why, for example, would a brain state serve as a criterion for pain? Perhaps a brain state could serve as a criterion through a new convention. But then we would have to consider the reasons for inventing such a new convention.

Von Wright’s view is not as clear as the previous quotes may suggest. He writes: “[…] one can have pain without feeling it. But this surely is a marginal case… a person can…

¹¹⁰⁰ Von Wright, 1998, 57.
¹¹⁰¹ Antony, 1989, 186.
in marginal cases, be said to have pain without feeling it, but only provided he exhibits pain behaviour.”¹¹⁰² In von Wright’s case, as in Davidson’s, behavior is the main evidential basis for mental phenomena. However, I think we should be dissatisfied with the suggestion that one can have pain without feeling it. What would the expression “he has pain” mean if not that the person in question feels pain? What does it mean to say that Georg has pain but does not feel it? If it means that Georg exhibits pain behavior, then the form of behaviorism in question seems to be very strong. Von Wright claims, puzzlingly in my view, that the feeling of pain is not necessarily required for Georg to have pain if Georg exhibits pain behavior. Yet, he comments that: “I differ from behaviourism… in that I cannot accept an identification of the mental with the state of affairs in the material world.”¹¹⁰³ I think we should say that for the observer behavior is the main evidence on the basis of which he attributes pain to a subject, but that in the first person case the main evidence is the feeling of pain. I believe that von Wright’s final views about consciousness, which were described in section 2.5.1, come close to this kind of position. Any sensations of mine of which I am not aware simply do not exist. It would certainly sound strange to say: “I have a tremendous pain but I don’t feel it.” Could I attribute the pain to myself on the basis of my pain behavior? Would there be such behavior, e.g. groaning, crying etc, without the feeling of pain?

The exact nature of pain as a mental phenomenon is not clear, but it seems to me that there is a strong mental component, namely the feeling related to pain, and perhaps it would not be inaccurate to say that this is pain. People can suppress their pain by focusing on it or they may increase the severity of pain by paying attention to it. Is pain to be identified with the physical damage to the body or is it to be identified with the feeling which results from this damage? Or is the feeling the subjective aspect of the objective bodily damage? It seems strange to say that the damage is identical with the feeling; the physical damage may be in the foot whereas the feeling of pain is not in the foot. Perhaps all answers about the nature of pain are inconclusive and unsatisfying. The phenomenon of phantom pain, the feeling of pain in an amputated limb, could suggest that pain is not to be identified with damage to the body. But where is phantom pain? It cannot be in the amputated leg; yet it surely feels as if it is there. Von Wright notes that if pain is understood as something “which is there in the body” and of which I am not aware, then pain is nothing mental, it is more like something belonging to the material world. The subjective, mental, aspect of the pain, which I see as an essential feature of pain is, according to von Wright, a “marginal case of the real”. This claim relates in an

¹¹⁰² Von Wright, 1998, 169, my emphasis.
interesting way to von Wright’s neutral monism. A pure mental state without any physical manifestation is a counterfactual construction; to say that a person has pain but does not exhibit pain behavior is to attribute to him a (mental) state in which a person is when he exhibits certain kind of behavior which would make us to conclude that he is in pain. The “impenetrable subjectivity” of the mental is a marginal case because a statement about the existence of a pure mental state has to be construed counterfactually in terms which refer to physical phenomena. But the same is true of physical phenomena; their status as totally mind-independent and objective is also a marginal case of the real, because a statement of their existence must be constructed counterfactually in terms of sensations. This is the conceptual interdependence of mind and matter.

The criteria of the mental are behavioral. That the criteria of the mental are behavioral and therefore independent from the neurological means that mental phenomena have epistemic priority in relation to their neural equivalents. The relationship of the three levels, mental, behavioral and neural, can be illustrated in the following way:

\[
\text{Causal relation} \\
\text{Macro-behavior (B)} \quad \rightarrow \quad \text{Mental (M)} \quad \rightarrow \quad \text{Neural (N)}
\]

\[
\text{Semantic relation} \quad \text{Epistemic relation}
\]

Behavior B, for example pain behavior, means for us that a subject S is experiencing a mental state of pain, M. The attribution of M is ultimately grounded on the semantic relationship between B and M, and the very possibility of such an attribution is based on this relation. This, in essential respects, is the point emphasized also by Davidson when he claims that the only way to find out whether a creature believes, desires or wants, is to interpret its behavior through familiar methods.

Once we have identified the presence of M, we can consider what the causal relation between its behavioral manifestations and neural states is. What are the neural causes of B? In order to answer this question, we must first be sure that the subject S is experiencing M. M has thus epistemic priority with respect to the neural, and the existence of M can be decided only from the presence of B. The mental could be said therefore to have a kind of double-priority with respect to the neural. The presence of M can be judged only by studying the behavior of a subject, and this study is essentially something which is guided by our
understanding of how to make best sense of physical phenomenon B. It is thus the active, mental contribution of the interpreter which answers the question of what B can be taken to mean. Pain, as a mental phenomenon, cannot be attributed to a subject who, for example, is faking pain. Once the presence of M is established, the question about its relation to N can be raised. But the question about the relation of M and N is always secondary with respect to the question about the relation of M and B.

I think von Wright’s observations have enormous importance. The fact that mental phenomena are epistemically prior to their neural counterparts demolishes the philosophical motivation for psycho-physical identity theories. In the case of ordinary psychological concepts, we have established the existence of the referent of the concept by relying on behavioral evidence that is interpreted by the active interpreter. This evidence is not only a symptom of the mental phenomenon. It is a criterion which must be satisfied if we are to talk about the presence of the phenomenon at all. The reason for this is that the mental is, at least on a Davidsonian–von Wrightian view, conceptually construed by placing it in a certain kind of context. The possible mental–neural correlations are only derivative from the fact that the existence of the mental phenomenon has already been established on independent grounds. Even if correlations between the mental and the physical were to be found, the existence of such correlations would not be as basic as the fact that mental phenomena have epistemic priority with respect to their physical counterparts. When deciding whether a mental phenomenon is present, the existence of psycho-physical connections would always be secondary when compared to the ways that we usually come to know about the existence of the mental. This is to emphasize the distinction between symptoms and criteria.

The epistemic priority of the mental and the semantic priority of the behavioral suggest that there is a principled reason against the possibility of strict mental–physical relations. Even if it made sense to speak of these kinds of relations, their existence would depend on the fact that the existence of the mental would have been already established on independent grounds. Von Wright claims that the very possibility of correlations between the mental and the physical requires that we can assert the presence of one without the other. He notes: “[...] unless one already knows what it is to hear, one cannot make discoveries about the neural basis of hearing. Or one would not know that these discoveries are of relevance to hearing.” More precisely: “In order for us to correlate the [mental and physical] phenomena the concepts must be there and have at least so much stability that classification

1104 Von Wright, 1998, 92. This point is stressed also by Malcolm, 1971.
under them is, in most cases, unproblematic.”

If we did not know how to identify phenomena falling under mental concepts, we could not establish their correlation with other phenomena either. This view is based on the already emphasized obvious fact that we have meaningful mental concepts which we use fluently. In answering the question “Is a person in a mental state X?”, we rely on a very different kind of criterion than what we would take to be relevant in answering the question “Is a person in a brain state Y?”, because mental states have properties which the physical states lack. By considering examples of mental concepts and physical concepts, we see that the identification criteria for them are essentially different.

Consider the case of a person who suffers from cancer. The reason why we conclude that she has cancer is not her subjective feeling about her situation. If she just feels that she has cancer, we would be suspicious no matter how strong the feeling is. She may complain that she has headache, nausea, and that she is tired. She acts accordingly. However, these symptoms do not warrant the conclusion that the person has cancer; a blood test or an X-ray or some kind of medical examination is required, for the behavior cannot be conclusive. But consider now a person who is depressed. According to one understanding of what “being depressed” means, we would not say that a blood test or a brain scan would warrant the conclusion that the person is depressed. In this case, as in the case of pain, the subjective feeling of the person as well as his statements and behavior are relevant. It would be strange to say to a person who seems to be very happy and acts accordingly: “Unfortunately the picture of your brain tells us that you are very depressed.”

Is neurological evidence irrelevant when considering whether mental states are present? Knowledge of the subject’s neurology could certainly be relevant in deciding whether, for example, a subject heard a sound. It is plausible to think that the neural could play some part in questions about the mental. But as we have already seen, supposing that the neural could play a decisive evidential role requires that we have established a mental–physical correlation, and this in turn requires that we have already established facts about the mental independently. These facts would have their criteria in non-neural phenomena. From this it follows that the neural criteria would necessarily be secondary in relation to the criteria which we take to be constitutive for the mental phenomenon. This means that the neural level could not function as a criterion but only as evidence or symptoms for the mental. As von Wright suggests:

---

1105 Von Wright, 1998, 93.
1106 I am here supposing that “being depressed” counts as mental phenomenon. That is, a feeling of depression is necessary if the subject is correctly described as being depressed.
If the correspondence [between the mental and the physical] is well established, the neural state in question may be regarded as a reliable sign or symptom of the existence of the reason. As long as the correlation remains a scientific hypothesis, the neural state fulfils this role of a symptom.\(^{1107}\)

There is therefore no principled reason why knowledge of the neural could not be helpful when deciding questions about the mental.

We conclude that that neurological evidence could have relevance in the questions about the mental. We can admit this, and yet insist that we must first know whether a person is in a mental state before we can investigate what physical state of her brain corresponds to this state. The mental level is epistemologically privileged in a fundamental sense because the initial observations of the presence of a mental state must be made without any knowledge of the neural facts. From this, it follows that the role which we allow for the neurological evidence depends on the question of how well it fits with the already established mental facts. This strengthens Davidson’s claim that strict correlations between two domains depend on the question of how well the predicates can be related to each other. We have first established the connections between mental predicates to each other, and in establishing these connections various principles, like the principles of normativity and holism, have been applied. The question of how well neural predicates fit into this story is a secondary question.

If we once more consider the case of the scientists who study a subject through their computer interfaces, what should we say? What would their situation be as compared to the group that explains and predicts the behavior of a subject with familiar methods by studying his linguistic and non-linguistic behavior in a normal environment? I think we have to conclude that we would have no reason to believe that the group studying just the brain could ever be sure in their predictions and explanations. This observation applies, of course, also to those who study the subject in familiar ways. However, they would be in a privileged position in the following sense: the scientists could use the data from the computer as evidence for their views, but this would not be conclusive evidence and they would therefore be in a worse position than those who could study the individual in familiar ways. Familiar ways would not be “conclusive” either, but given what we mean by mental concepts, if the subject fulfilled the criteria required for the attribution of mental phenomena, there would be no firmer ground for attributing these phenomena to him as the ordinary way. The scientists would have to see what the subject does in order to check whether their explanations and predictions are correct.

Although the situation of the scientists consulting their computers would be hopeless in the sense that they would only have secondary evidence at their disposal, we could still

\(^{1107}\) Von Wright, 1998, 37.
suppose that they might have established correlations between mental and physical states. Indeed, they would make their predictions, and after checking them through familiar means, they could conclude that the established correlations held, at least with very high probability. Should we now say that they have formed psychophysical *identities*? Should we think that the correlations established could be used for the formulation of laws? In answering this question, we should consider Davidson’s somewhat cryptic remarks such as the following.

The thesis is... that the mental is nomologically irreducible: there may be true general statements relating the mental and the physical, statements that have the logical form of a law; but they are not lawlike (in a strong sense to be described). If by absurdly remote chance we were to stumble on a nonstochastic true psychophysical generalization, we would have no reason to believe it more than roughly true.  

Or,

[...] if we were to find an open sentence couched in behavioural terms and exactly coextensive with some mental predicate, *nothing could reasonably persuade that we had found it*. We know too much about thought and behavior to trust exact and universal statements linking them.  

Or,

As long as it is behaviour and not something else we want to explain and describe, we must warp the evidence... Standing ready, as we must, to adjust psychological terms to one set of standards and physical terms to another, we know that we cannot insist on a sharp and law-like connection between them....

One could well wonder what it means to say that we have no reason to believe that psychophysical generalizations are more than roughly true, or that we know too much about thought and behavior to trust exact and universal statements linking them. What does it mean to say that in the case of the mental which manifests itself in behavior, it may be that we are warping the available evidence? What does it mean that the evidence has to be twisted? The way I read these important statements is that they are just specifications of the general view that we cannot allow the possibility of strict psychophysical laws. Since the argument is that we cannot allow the possibility of psychophysical laws, a “pragmatic” justification lies behind Davidson’s reasoning. I think the claims above make sense if the primacy of interpretation in establishing facts about the mental is taken seriously. Both von Wright and Davidson insist that an ascription of a mental state is always open to revision over the course of time. There is always a possibility for re-interpretation which is governed by constitutive principles of the

---

1108 Davidson, 1970, 217, my emphasis.
1109 Davidson, 1970, 217, my emphasis.
1110 Davidson, 1974b, 239.
mental. We cannot allow the possibility that certain physical facts would conclusively warrant our claims about the presence of mental states. If correlations between the mental and the physical were established, they would be mere empirical generalizations and not even close to strict laws in Davidson’s sense of the term.  

Both Davidson and von Wright emphasize the priority of behavior over neural processes in establishing the presence of mental phenomena. But, according to von Wright, to ascribe mental states to a creature is not to ascribe a specific behavior or sets of behavior to it. Why? Because:

The criteria of... mental phenomena have the following peculiarity: No finite combination of them is (logically) sufficient to establish the presence of the phenomenon.... Moreover, no finite combination of behavioural criteria is (logically) necessary for the presence of the sensation.... For this reason I shall say that the attribution of... mental phenomenon to a being has a residue of meaning which is not captured by any enumeration of the behavioural reactions which are constitutive of its meaning.  

Because of the residue of meaning, it makes no sense to say that a mental phenomenon would be identical to a set of physical processes or behavior.  

Both von Wright and Davidson have a behaviorist approach to the mental which is not reductive. The view that there is no necessary or sufficient physical application condition for a given mental phenomenon leads to the conclusion that physical properties are the wrong kind of thing for us to draw mental inductive inferences from. This reaffirms Davidson’s claim that mental and physical predicates or properties are not made for each other. It seems to me that what von Wright argues clarifies what Davidson means when he writes:

When we use the concepts of belief, desire, and the rest, we must stand prepared, as the evidence accumulates, to adjust our theory in the light of considerations of overall cogency: the constitutive ideal of rationality partly controls each phase in the evolution of what must be an evolving theory. An arbitrary choice of translation scheme would preclude such opportunistic tempering of theory; put differently, a right arbitrary choice of translation manual would be of a manual acceptable in the light of all possible evidence, and this is a choice we cannot make.  

Openness of interpretation forces us to be ready to “warp the evidence” in terms of the principles of interpretation. Because the mental pattern of a person must obey the constitutive ideal of rationality, there must be a possibility for “opportunistic tempering”, that is, a

---

1111 As Davidson (1964, 47) says: “what we want as a law of correspondence is not an accidental correspondence, but one which, we have reason to believe, will apply to cases which we have not yet examined.”  
1112 Von Wright, 1998, 146.  
1113 The residue of meaning is what also, according to von Wright, accounts for the idea that mental states are “hidden” or “private”. I showed in section 2.6 that, according to von Wright, the “location of the mental” is in the residue of meaning.  
1114 Davidson, 1970, 222.
possibility for an adjustment *according to this ideal*. We can suppose that we could translate our mental vocabulary into a physical vocabulary. This kind of translation would be an *arbitrary* choice from the perspective of the mental, because we could never be in a position where all the mental facts would be available. We cannot know in advance what mental facts are relevant for a specific interpretation. No finite combination of the criteria for a mental phenomenon is sufficient or necessary for the presence of the phenomenon. No set of behavior B or set of neural phenomena N could conclusively justify our attribution of M to a subject.

Following Davidson and von Wright, I have argued that the mental is epistemically prior to the neural. Many find this reasoning unsatisfying. Antony, among others, thinks that, *contra* von Wright, there must be a truth about mentality which exists independently of what anybody thinks about these matters. But naturalists have not given a reason why we should favor this kind of conception of mentality. The only justification for the view is the question-begging claim that anything real *must* be mind-independent. But consider, for example, the phenomenon of jealousy. We observe Donald acting in a way which any reasonable person would take to be a “certain” sign of jealousy. When asked about the situation Donald denies that he is jealous. How are we to decide who is right? Suppose that we can *convert* Donald to realizing that he is jealous by pointing his attention to the obvious signs that we observed. Then Donald could actually admit that, also to his own surprise, he is jealous, or he could have a counter-argument which would force us to re-consider our initial opinion. This debate could continue until a consensus was reached or perhaps the case would stay unresolved forever. What is now, after the consensus, *the truth* about the situation and how has it been reached? Was Donald really in a state of jealousy? Or is it more reasonable to conclude that the state of jealousy is a shady state which may be there at one moment and be absent at another moment? If mental states are in this sense social objects, it is by no means clear that we should see their nature in a similar way to how we see the nature of objects or events, which seem to be mind-independent in a clearer sense.

However, there seems to be *some* relation between the mental and the physical, even if we think that the relation cannot be expressed in terms of strict laws. Here is a von Wright’s conclusion that many contemporary philosophers would find flawed: “What the neurological equivalent [of a mental state] is, is not philosophically interesting, nor even whether there is an equivalent. The “philosophical” question is whether there *must* be an equivalent”.

Von Wright, 1998, 92.
Wright thinks that the interesting philosophical question about psycho-physical parallelism is whether it must be true. He concludes that, in some sense, its truth must be a contingent, because the knowledge we have of the mind–brain correspondence is contingent knowledge. As von Wright notes: “Man was familiar with mental phenomena... long before... one knew that there was such a thing as a nervous system.”\textsuperscript{1116} This being the case: “Is it not thinkable that everything in the world of the mind, including intentionality of behaviour, went on in the way it does even if the brain and nervous system did not even exist?”\textsuperscript{1117} We should conclude that that the relation between behavior and mental states is conceptually independent of anything that happens in the nervous system. The connection between behavior and the mental is essential whereas the fact that there is a nervous system linking the two is accidental. At least on the Davidsonian–von Wrightian account, one important task of philosophy of mind is to analyze the language of the mental and this language is conceptually autonomous with respect to anything that takes place in the brain. I think that von Wright makes a daring claim when he argues that “[...] all philosophical problems about the body–mind relationship, about ‘the inner and the outer’, and about consciousness can be satisfactorily coped [with] while ignoring the fact that human beings... have a brain and nervous system.”\textsuperscript{1118} This is a very challenging claim. Von Wright concludes:

If my argument is correct, it has a remarkable consequence for the ‘place of the brain’ in the body–mind problem-complex. Since the mental has epistemic priority in relation to the neural... the mental... cannot be ‘reduced’ to the neural... In establishing the neural correlates we rely on the psychological phenomena as given to us in experience. From the psychological point of view, the very existence of a nervous system, ‘mediating’ between the world of the body and the world of the mind is, logically speaking, an accident.\textsuperscript{1119}

I think that the epistemic priority of the mental in relation to the neural is a crucially important point, which should be taken seriously in the modern philosophy of mind. It is a similar kind of philosophically deep conclusion as von Wright’s observation about the mutual dependence of mind and matter. But how could we take seriously the claim that philosophical problems about the mind can be coped with while ignoring the fact that human beings have brains? Nothing would sound more outrageous from a naturalistic point of view. Von Wright himself thought that his statement was a bad and clumsy expression which nevertheless expressed a valuable philosophical insight. I think that the most important thing to notice is von Wright’s

\textsuperscript{1116} Von Wright, 1998, 134.
\textsuperscript{1117} Ibid.
\textsuperscript{1118} Von Wright, UPg, 4. Von Wright notes that this is the road which Wittgenstein tried to take and it makes his philosophy so uncongenial to many. I have suggested that the same is true of von Wright and Davidson.
\textsuperscript{1119} Von Wright, UPf, 4.
claim that all *philosophical* problems about the body–mind relationship can be clarified without taking into consideration the fact that humans have brains. I am not sure whether this is true of all philosophical problems about the body–mind relationship; perhaps the question of whether mind is identical with the brain is a philosophical problem. Because of his specific conception of what a philosophical problem is, von Wright would deny this. I believe the claim makes sense if von Wright’s view about the nature of a philosophical problem is accepted. On the other hand, it seems to me that a philosopher like von Wright, who is committed to O-physicalism, must find *some* substantial place for the brain in the mind–body problem. Von Wright notes that making the claim sound plausible is itself a formidable philosophical task, because the suggestion goes against strong philosophical tendencies and temptations which are driven by trends in contemporary science. The strong philosophical tendency is, of course, the naturalistic approach in the philosophy of mind. Von Wright wanted to defy the scientific trends in the name of philosophy, although he recognized that this approach is: “...a daring adventure which may fail and not succeed.”  

This remark shows that, as I have suggested, von Wright wants to challenge what he sees as the prevalent unhealthy scientific philosophy in the name of pure philosophy. When the claim is seen from this perspective, it invites us to consider again and again the question of what the task of a philosopher is in the explication of the mind–body problem.

After suggesting that facts about the brain can be ignored in the philosophy of mind, von Wright notes that someone could argue that it is a demand of “scientific intelligibility” that there must be a neural equivalent for a mental state. But, it seems to me that the discussion emphasizing the epistemic priority of the mental in relation to the neural shows precisely that this scientific claim about the existence of equivalencies should not be confused with the philosophical question of whether such equivalencies are necessary. Let us suppose that we used the mental vocabulary to describe the behavior of creatures which turned out to be creatures without nervous systems. Would we, after this scientific finding, say that our mental vocabulary does not apply to these creatures? If the answer is affirmative, what was our reason to use that vocabulary in the first place? How is its successful usage explained before any scientific discoveries were made? How, in the past, would we have explained its success if we were completely ignorant about the brain?

Although von Wright thinks that it is hopeless to try to argue that psycho-physical parallelism is necessarily true, he notes that one can still *postulate* its truth and use it as a

---

1120 Von Wright, UPg, 5.
heuristic device for encouraging brain-research. It is possible to stick to this postulate, as naturalists do, and try to find out what kind of connections (if any) there are between mental and physical states. Indeed, as von Wright notes: “[…] the details of the [brain-mind] correspondence can only be found by scientific, neurophysiological research.”\textsuperscript{1121} If we stick to the postulate the correspondence between the mental and the physical may seem to be an \textit{a priori} feature of reality. In this respect von Wright sees the postulated correspondence as a “demand of reason” which we urge ourselves to employ. Likewise, Davidson does not exclude the possibility that we could find connections between the mental and the physical, and in this sense there is no reason to stop doing brain-research.\textsuperscript{1122} Nevertheless, because of his quite demanding understanding of what it takes for a law to be “strict”, he firmly claims that the mental and the physical cannot ever be paired in terms of strict laws.

3.3 The irreducibility of the mental

We should recall, from the beginning of this chapter, what Davidson’s claim about mental concepts is: “[…] they cannot be defined in the vocabularies of the natural sciences, nor are there empirical laws linking them with physical phenomena in such a way as to make them disposable.”\textsuperscript{1123} In this section I will consider the question of what it means to say that mental concepts cannot be defined in the vocabulary of the natural sciences, and why this is the case. The claim that there are no empirical laws linking the mental and the physical was discussed in section 3.1 and therefore, as long as we accept Davidson’s arguments, the possibility of nomological reduction is already precluded. What I will consider here is the question what the impossibility of definitional reduction means and what are the reasons for it.

The thesis of the irreducibility of the mental can be approached from various directions, and it is closely connected to the thesis of the anomalism of the mental. On the one hand there is the question of whether mental \textit{concepts} are irreducible to physical concepts. Given that, according to Davidson, the mental is only a conceptual category, this kind of conceptual irreducibility could be taken to be all the irreducibility there is. However, we should also consider in what sense the mental can be \textit{ontologically} irreducible. Although Davidson’s philosophy of mind is meant to be ontologically reductive, there is also a clear sense in which it is not. If one chooses to defend the irreducibility of the mental, one may lose

\textsuperscript{1121} Von Wright, 1998, 134.
\textsuperscript{1122} Although, as I showed in section 2.2, Davidson’s last word on the subject seems to be that we have no reason to believe that token-identities could be found.
\textsuperscript{1123} Davidson, 1999o, 599.
the physical status of the mental. The claim that all events or states or properties are physical is obscure if it turns out that we cannot identify mental events or states or properties with their physical counterparts at the “lower” ontological level. The problems of type—type and token identities have already been discussed in chapter two. It could be said that by emphasizing the irreducibility of the mental we are, step by step, questioning whether the basic claims of physicalism make sense. It is unclear whether this was Davidson’s purpose since, as we have seen, he wanted to defend O-physicalism. On the other hand, arguments defending the irreducibility of the mental undermine the truth of physicalism. It could be argued, though, that the irreducibility of the mental only undermines the truth of physicalism as usually understood, and that it undermines the importance of traditional arguments in favor of physicalism. However, I think that we should consider what is left of physicalism if the principled irreducibility of the mental is allowed. I leave this question aside for the moment, consider the arguments for the irreducibility of the mental, and then suggest what follows for physicalism.

There are three reasons for the irreducibility of the mental that can be found in Davidson’s views. They are:

1) Normativity, rationality and the subjective nature of thoughts
2) The dispositional character of mental concepts
3) Externalism

Before turning to the arguments, something should be said about the kind of reduction that Davidson is opposing. The concept of “reduction” is sometimes unclear in the philosophical discussion, and problems occur because “reductivists” and non-reductivists” seem to be talking about different things. This is a similar kind of problem to the one we encountered in the discussion about the anomalism of the mental. Philosophers talk past each other because they have different views of how strict the psycho-physical laws should be, or of what kind of identities a philosopher should establish. It is not always clear what a philosopher is claiming when he is arguing for reduction; likewise when a philosopher is arguing against it.

---

1124 As has been stressed, among others, by Kim, 1989, 1998.
1125 For a general discussion about reduction see, for example, Beckermann, Flohr, Kim (eds.), 1992 and Hohwy and Kallstrup (eds.), 2008.
3.3.1 What does a reduction of the mental to the physical mean?

If we go through the philosophical literature on reduction for the last twenty years, it becomes obvious that it is not clear what it means to say that mind could be reduced to matter – or that this cannot be done. Most philosophers do not believe in substance dualism, and therefore for many the appeal of the old mind–body problem has disappeared. New questions around the old mind–body problem are appearing, however. Most modern philosophers think that mind is a “physical device”– the brain – and many of the disputes concern the question of what this statement means. We may be quite convinced that the mind is indeed a “physical thing”, but the question is what puts us in a position to say this. What are the reasons to argue that the mind is a physical thing? What are the reasons to argue that it is not, and what could it mean to say that the mind is something non-physical? I think that despite the consensus around O-physicalism, these are pressing questions.

With respect to these problems, a lot depends on the question of whether mind can be actually reduced to matter. This question may turn out to be crucial also in deciding whether one can be a physicalist with a good conscience. The question of reducibility is often described in a way which leaves only two options for a physicalist: emergence or reduction. If there is no explanation for how mental states reduce to physical states, it has to be accepted that mental states are in some sense emergent, which in turn implies that their physical status is, at the very least, obscure. If mental phenomena turn out to be emergent phenomena, then the status of physicalism is in jeopardy. Crane, for example, has claimed that whether an interesting version of non-reductive physicalism is possible depends on whether conceptual reduction is possible. He thinks, contrary to Davidson, that this is an empirical question, but concludes that since the question is still open, the question about the truth of physicalism is open as well.1126 I have emphasized that as long as the mental remains irreducible, we have no reason to believe that physicalism is true. If, contra Crane, we accept the Davidsonian view that the possibility of conceptual reduction is not an empirical question, then the Davidsonian irreducibility, if correct, would seem to show that physicalism cannot be true.

Let us suppose that we want to show that mental states reduce to physical states. How could this be done? What would assure us that we have succeeded? A fairly standard view is that a science or its vocabulary might reduce to another through: a) ontological reduction, b) nomological reduction, or c) semantic reduction. All these types of reduction are

1126 Crane, 2008.
somewhat vague and overlap each other. It is not always clear what a reductionist wants. I
shall nevertheless try to describe what a reductionist attempts to accomplish and consider
whether this is something prevented by Davidson’s argument. The purpose of this
consideration is to place Davidson’s views in a contemporary perspective and to see whether
his arguments have any relevance any longer to those who defend the autonomy of the
mental. One of Davidson’s insights was that – allegedly – one can be a naturalist without
being a reductive physicalist. Non-reductive physicalism in general, and Davidson’s version
in particular, endorses ontological reduction but stands against nomological or definitional
reduction. Davidson notes that the concept of reduction, which he accepts, and actually
defends, is ontological reduction and that it is “trivially obvious” that this weak understanding
of reduction does not entail definitional or nomological reduction. Davidson complains that
philosophers often reason from ontological to a stronger form of reduction or simply conflate
these two. Whereas Davidson strongly endorses ontological reduction, he just as strongly
argues against the possibility of nomological and semantic reduction. Indeed, as it is by now
familiar, Davidson tries to establish the truth of a physicalistic monism from the fact that any
semantic and nomological reduction between the mental and the physical must fail.

Weak ontological reduction is not what a reductionist really wants. He thinks that
ontological reduction can be achieved only through semantic or nomological reduction. A
contemporary reductionist cannot be satisfied with ontological reduction if it is achieved
through a conceptual a priori argument. A weak ontological reduction either demands other
forms of reduction or entails them. It seems to me that two reductionist lines of thought can
be detected. On the one hand, ontological physicalism is the goal and its truth can be fully
established only if it can be shown how higher level phenomena reduce to those at the lower
levels. From this perspective, non-reductive physicalism promises too much when it claims
that physicalism and non-reductivism make a coherent combination. On the other hand, the
truth of ontological physicalism can be treated as an assumption, and then the argument is
that, since everything is physical, a reduction of higher level phenomena must be possible, at
least in principle. No matter which of these options is chosen, reductions turn out to be vitally
important for a reductive naturalist.

Semantic reduction is one way in which the argument for physicalism has been
formulated. In the first half of the 20th century, the suggestion that any sentence S of any
science could be translated into a sentence of a physical language was attractive to many
philosophers. Logical behaviorism claimed that statements made in our mental vocabulary
can be analyzed into statements in the vocabulary of physical behavior, was a special case of
this general view. Davidson does not spend too much time in his rejection of semantic reductionism. In “Mental Events” he briefly notes: “Why are we willing... to abandon the attempt to give explicit definitions of mental concepts in terms of behavioural ones?... [it is] because we are persuaded, as we are in the case of so many other forms of definitional reductionism... that there is system in the failures.”

He then emphasizes that when giving non-mental conditions for mental events we always find the need for an additional condition that is mental in character. In another context, Davidson refers to Frege’s argument that intensional sentences seem to resist analysis in terms of extensional sentences, which, given Davidson’s views about the mental, means that mental concepts cannot be given non-mental definitions. On the whole, Davidson concludes that there is “overwhelming evidence” against the view that mental concepts could be defined in terms of physical concepts. This being the case, in his philosophy of mind Davidson does not present an explicit argument against semantic reductionism.

Let us accept Davidson’s view that semantic reductionism does not look promising and turn to the question of whether a nomological reduction of the mental to the physical is possible. We have already seen why Davidson thinks that this kind of reduction is also impossible. There is nevertheless a question as to whether Davidson’s argument really blocks the possibility of nomological connections between the mental and the physical, and even if it does, whether there is some other sense of nomological reduction which his arguments do not touch. Successful reductions in science occur between theories. In the case of mental–physical reductions, we should not therefore straightforwardly ask whether mental states can be reduced to physical states but whether a theory, or vocabulary, which refers to these states, reduces to a theory which does not refer to them.

The conception of reduction that long dominated the discussion about the possibility of mind–body reduction was the model suggested by Ernest Nagel. Reduction of a theory $T_r$ to $T$ is achieved when $T_r$ is deducible from $T$. In many cases the deduction of $T_r$ from $T$ is possible only if bridge-laws that connect the expressions between theories are formulated. Bridge-laws are meant to be empirical laws stating biconditionals; thus in the supposed case of mental to physical reduction, there could exist a bridge-law of the form “for any object $x$, $x$ is in $M$ iff $x$ is in $P$” where $M$ refers to a mental property and $P$ refers to a physical property. Given this view about reduction, the consequences for the mind–brain identity theory are straightforward. Mental phenomena are identical with physical states iff psychology is

---

1127 Davidson, 1970, 217.
reducible to, say, cognitive neurobiology. Mental phenomena are identical with physical states, if there exist bridge laws that connect each mental predicate with a suitable neurobiological predicate in a way that the laws of psychology can be deduced from neurobiology with the aid of these bridge-laws. This view about reduction suggests that the existence of psycho-physical bridge-laws is necessary for any identity theory to be true. As we have seen, Davidson denies the possibility of these kinds of mental–physical bridge-laws, and therefore denies the possibility that mental states could be Nagel-reducible to physical states.

In the case of Nagel-reduction as applied to possible mental–physical reduction, the interesting question is whether bridge-laws provide identities. A fairly standard interpretation is that in Nagel’s model, bridge-laws are indeed meant to provide identities between two predicates or properties. But what kind of identities are the consequent identities? If a bridge law connects M and P, is it the case that anything that is true of M is true of P, and that anything that can be explained by M can be explained by P? Nagel was not sure whether the nature of bridge-laws should be seen as being “conventional” or “factual”. The question about the nature of bridge-laws is interesting, because if the relation is meant to be factual, then the question of whether there can be psycho-physical bridge laws seems to depend, partly, on whether sufficiently fine-grained correlations between mental and physical states can be found. This is an empirical question, and the question about the possibility of bridge laws was usually treated as such when philosophers first started to defend type–type identity theory.

Nagel-type reductionism, which is essentially the type of reduction that Davidson’s arguments for the anomalism of the mental are meant to refute, fell quickly into disgrace. The main philosophical reasons why this kind of classical model of reduction was something that could not secure a mind–body identity theory were: a) the obscure status of bridge-laws, b) functionalism, and c) anomalous monism. Whereas it was first thought that type-identity theories could benefit a lot from a general view of how reduction in science proceeds, the philosophical discussions concerning the above-mentioned topics soon started to create doubts against this optimism. Another reason, which was then used to formulate different strategies of reduction, was the realization, based partly on historical evidence, that reductions in science do not usually happen in the way described by Nagel. It was noted that in the

---

1130 A classical example is Smart, 1959, 1963.
1131 See for example, Schaffner, 1967 and Hooker, 1981. The general work of Kuhn and Feyerabend was also relevant for this realization. On the view put forward by Hooker a new way to think about reduction was formed which is continued, for example, by the Churchlands and Bickle.
case of two theories $T_r$ and $T$, it is not usually the case that the original $T_r$ can be deduced from $T$ without any modification of the original theory. This being the case, it is not actually $T_r$ that is being deduced from $T$, but a corrected version of it. Moreover, this being the case, it is unclear that bridge-laws would be necessary for a successful reduction.

Let us suppose that theory $T_r$ is in some important respects a false theory and that certain entities or properties of the theory do not exist. Then there is no reason to try to correlate these entities with anything in the reducing theory. Since reductions usually involve revision and modification, and that is one of their goals, we should conclude, according to Patricia Churchland, that “whenever the corrections to the old theory are anything more than relatively minor, it is always tendentious to claim that phenomena in the old theory are to be identified with phenomena in the reducing theory.”\footnote{Churchland, 1986, 282.}

In cases where the modifications to the old theory are notable, bridge-laws play no role. Paul Churchland claims that cross-theoretical identity claims are not part of a proper reduction and that they are not essential to the function that a reduction performs. In the cases where the reduction is not “smooth”, bridge-laws, or correspondence rules, as Paul Churchland calls them, can be treated as mere ordered pairs of expressions. The important point is this: reduction is possible, and often proceeds without strong bridge laws that should be treated as identity claims. This being the case, there is an increasing tendency to argue that bridge-laws are not necessary for reduction. After discussing certain problems relating to the nature of bridge-laws and concluding that reductions based on universal biconditionals are extremely rare or non-existent in the sciences, Kim concludes: “...the question whether or not mentality is Nagel-reducible via bridge laws to the physical cannot be a significant metaphysical issue. Nor can a refutation of reductionism that is premised on Nagelian reduction be considered a significant philosophical contribution.”\footnote{Kim, 1998, 97.}

According to Kim, Davidson’s argument refuting reductionism based on bridge-laws would thus not count as a significant philosophical contribution.

It appears that definitional reduction of logical behaviorism does not seem to be a promising option for a reductionist. Nomological reduction, understood in the Nagelian way, is not a viable option either because, in addition to the philosophical arguments against it, philosophers have concluded that there are cases of successful reductions which do not necessarily involve bridge-laws in the Nagelian sense. Since reductions can take place without bridge-laws, a reductionist wants to find out whether mental reality is also reducible to the physical through these kinds of reductions. Davidson claimed that his physicalistic view does
not entail reducibility *through law or definition*. We have seen that reducibility through definition is implausible, although Davidson also admits that there is no proof against this possibility, only “overwhelming evidence.” We have also seen that reducibility through law does not seem plausible because reductions take place without the kinds of laws that Davidson has in mind. Although Davidson does not explicitly mention Nagelian reduction as his target, it is quite obvious that this is indeed the form of reduction that he takes himself to be refuting\textsuperscript{1134}. The purpose of the anomalism of the mental is to rule out reduction by way of *strict bridging laws*, laws which connect mental properties with physical properties, as Davidson notes. What is interesting and above else challenging to a Davidsonian position is the possibility of reduction of theory $T_r$ to $T$ *without* bridge-laws. As we have seen, this possibility is discussed and found wanting, for example, by the Churchlands and many other contemporary philosophers.

**3.3.1.1 New wave reductionism: A challenge for Davidson’s position**

The possibility of reduction without bridge-laws is raised as a challenge for Davidson explicitly by neurophilosopher John Bickle, who has become known for his model of “new wave” reductionism\textsuperscript{1135}. Bickle’s purpose is to keep identity theory “well and alive”, and he claims to be a “full-blooded physicalist”. These commitments require that Bickle defuse the Davidsonian challenge to identity theory. Without going into the details of Bickle’s theory, we should note that, according to him: “Davidson’s challenge, based upon the impossibility of psychophysical (bridge) laws, is entirely without force”\textsuperscript{1136} since on the new wave model, the connecting principles between two vocabularies or theories are *not* required to effect the derivation of $T_r$ from $T$. Bickle notes: “There are no disparate vocabularies to connect between premises... and conclusion.”\textsuperscript{1137} This is a problem for Davidson’s argument because it is based on the disparate commitments of the mental and physical vocabularies. Bickle concludes: “If the deductive part of reduction has no gap to bridge between the language or the ontology of premise and conclusion, then the nonexistence of lawlike connections between reduced and reducing concepts or kinds is of no consequence.”\textsuperscript{1138}

\textsuperscript{1134} For this interpretation see for example Mclaughlin, 1985 and Kim, 2003.
\textsuperscript{1136} Bickle, 1998, 108. See also Bickle, 1992.
\textsuperscript{1137} Ibid.
\textsuperscript{1138} Ibid.
Bickle claims that a Davidsonian argument against the possibility of psychophysical bridge-laws is no challenge to a conception of reduction which differs in essential respects from the Nagel-type reduction. Bickle thinks it is clear that Davidson offered his view as an alternative to a reductionism based on the Nagelian account of inter-theoretic reduction, and against this kind of reduction it is indeed a plausible challenge. However, if we recall Kim’s opinion that any “refutation of reductionism that is premised on Nagelian reduction [cannot] be considered a significant philosophical contribution”, it seems that we should conclude that Davidson’s argument against nomological reduction, even if successful, is not a significant contribution to the discussion about the possibility of mental–physical reduction. The situation seems to be this: Davidson is seen as an important figure behind non-reductive physicalism. Yet, he did not contribute to the refutation of semantic reductionism; its failure was already established. The Nagel-type reduction which Davidson challenges is of no interest to a reductionist in the philosophy of mind, since reductions do not usually occur in this way. According to Bickle, new-wave reductionism, the “most plausible” modern view of reduction, remains unchallenged by Davidson. So, we can ask whether Davidson’s views have any relevance for the contemporary discussion. Crane has commented that this is unlikely: “Davidson’s target is a very specific conception of what a conceptual reduction amounts to – essentially, reduction by bridge-principles… it is hard to see how the argument can be re-formulated to apply to the most plausible forms of conceptual reduction.” Unfortunately Crane does not tell us what he takes to be plausible forms of conceptual reduction. Whatever they are, I believe it is correct to say that the original argument for the anomalism of the mental cannot be re-formulated so that it would apply against all forms of conceptual reduction. But this does not mean that the arguments for the irreducibility of the mental could not be effective against other “plausible forms” of conceptual reduction. Crane seems not to fully recognize this, because he equates the view that there cannot be any conceptual reduction of psychology to physics wholly with the view that there cannot be psychophysical laws, that is, with Davidson’s argument for the anomalism of the mental.

I think we should agree with much of what Bickle says, namely that Davidson meant his argument against the classical form of reduction and that there is a plausible reading according to which the possibility of bridge-laws, and therefore of Nagel-type reduction, is compromised given Davidson’s arguments. If the possibility of bridge laws is blocked because of Davidson’s arguments, then I take this, contra Kim, to be a relevant contribution.

---

1139 Kim, 1998, 97.
1140 Crane, 2000, 80.
at least when seen in the intellectual atmosphere of the time when the argument was originally presented. It was, after all, AM and functionalism which were largely responsible for the fall of the type-identity theory. Given that the classical model of reduction was the most accepted option at the time of “Mental Events” and Davidson challenged it, it cannot be a straightforward criticism of Davidson’s view that it does not challenge or refute other models of reduction which have been developed later. In chapter two, I wanted to place AM in the historical context of its times. One reason for this was to consider how AM was relevant, given the intellectual convictions and agendas of the 1960s. One can only guess what counts as a “significant philosophical contribution” in Kim’s opinion; I think that the claim that a refutation of a reduction based on a Nagel-type reduction does not count as a significant contribution is made with the benefit of a hindsight.

Davidson’s argument can be seen as effective against classical forms of reduction. Does it also challenge the “new wave” approach? I think it is questionable whether the “new wave” approach is a feasible model of reduction. It is questionable whether Bickle’s model can actually avoid cross-theoretic identities. Endicott, for example, has argued that it can’t. The essential point in the new way of thinking about reduction is to see reductions as forming a continuum from theories that have been largely retained to theories which have been largely displaced. Between these extreme ends are vague cases where an old theory has been modified to varying degrees. Let us suppose that a psycho-neural reduction in fact falls at the retentive end of the continuum. Then, as Endicott notes: “[...] cross-theoretic property identities exist between reduced and reducing theories.” This being the case, “biconditional bridge laws exist between reduced and reducing theories”. Endicott concludes: “If a case falls at the retentive end of the new-wave continuum, then biconditional bridge laws exist between reduced and reducing theories.” On the other hand, if it is not the case that biconditional bridge laws exist between, say, psychological and physical theories, then it seems that psychological theories do not fall at the retentive end of the new-wave continuum. This being the case, we encounter the following pressing question. If mental

---

1141 As mentioned already, Bickle (1992) recognizes Davidson’s importance to the question concerning the possibility of bridge-laws. Also Kim (1989) acknowledges Davidson’s contribution, although he denies the importance of his refutation of Nagel-type reductionism.
1143 Endicott, 1998, 67. Consider Paul Churchland’s claim (1979, 83): “A successful reduction of the ideal sort described provides an excellent reason for asserting the relevant cross-theoretical identities, the best reason one can have”. Patricia Churchland argues (1986, 283): “If a reduction is smooth... then the matching of cohort denoting properties can proceed and identities can be claimed.”
1144 Ibid.
1145 Ibid.
states, or the science studying them, or the vocabulary which refers to them do not reduce to physical states or to physical science, then what should our conclusion be? If definitional and nomological reductions fail, why conclude that the irreducible entities or concepts are second-grade? Answers to these questions are bound to be connected to a given philosopher’s intuitions about the nature of such hopelessly broad matters as explanation, science, ontology, or the purpose of philosophy.

It is an open question whether Bickle or someone else can come up with a view of inter-theoretic reduction that can completely avoid inter-theoretic identities. But what is important is the following. As I noted, one way to defend Davidson’s argument against the possibility of mental–physical reduction is to claim that it was meant to discredit a Nagelian model of reduction and could not anticipate future models. This defense of Davidson could be continued by insisting that his view cannot be applied to forms of reduction which reject the view that reduction requires bridge-laws. Looking from the perspective of the 1960s and 1970s, Davidson could not anticipate what kinds of models of inter-theoretic reduction would be forthcoming, but I would claim that it was not his purpose to try to defuse views which reject the requirement of bridge-laws. However, this defense, although possibly correct and true to Davidson’s intentions, would be unsatisfactory in crucial ways. One defect of the defense is that Davidson’s argument is meant to be a philosophical argument. His arguments for the irreducibility of the mental are meant to be such that they cannot be refuted by any empirical evidence. This is obviously a very bold strategy and the aim is ambitious. If Davidson intends his claims about irreducibility to be irrefutable by any kind of reductive account, it is clear that he needs something more than the claim that there are no psychophysical bridge laws. His argument must apply also to modern models of reduction.

The problem involving reduction still troubling us is that it is not entirely clear what a successful reduction would achieve, how it should be done, and why a reduction would be desirable in the first place. Among reductionists there seems to be a widespread unity of opinion as to why reduction is a good thing. The main reason is that it gives both explanatory unification and ontological simplification. The latter may be achieved either through reduction and identification or elimination. Interestingly, Davidson would have no reason to praise reduction because of these benefits. The main theme in his philosophy of mind is that in understanding others we are forced to embrace conceptual dualism, and therefore the possibility of explanatory unification is blocked. As far as ontology is concerned, Davidson tries to show the truth of monism without reduction. To this view, an additional argument for monism which is based on the possibility of de facto reduction does not add anything, because
Davidson thinks he has already established the truth of monism on a priori philosophical grounds. One could conclude that Davidson does not see ontological reduction as a desirable thing, but this conclusion would also be incorrect. The correct way to see the situation, rather, is that ontology and reduction can be dissociated, and realizing this helps us to reduce the philosophical importance of each.\textsuperscript{1146} This view is at odds with positions like Bickle’s and the Churchlands’, who see ontology following reductions. As Bickle claims: “theory reduction first, ontological consequences second and dependent upon it....”\textsuperscript{1147} His view is straightforward:

\[
[...]
\text{‘new’ reductionists defend the view that the ontological status of beliefs, desires and the like is wholly dependent upon the nature of the reduction relationship obtaining (or failing to obtain) between intentional psychology and some future scientific successor.}\textsuperscript{1148}
\]

This kind of bold view is shared by most reductive naturalists and eliminativists. That questions of ontology must depend on the possibility of vocabulary reduction is a view which can be seen as a premise in many of the contemporary approaches in the philosophy of mind, and it is a view from which Davidson’s, not to mention von Wright’s or Wittgenstein’s approaches differ in crucial respects. But how could one argue that the view of von Wright and Davidson is correct? I think that on this issue there cannot be a convincing argument for either kind of position. The only thing we can claim is that in philosophy we should take seriously those things which actually have an essential role in our lives. To suggest that the ontological status of beliefs is wholly dependent on the question of how well these phenomena match with a future scientific theory is to blatantly ignore facts which are essential to our lives.

There are thus very different intuitions about the question of how the connection between reduction and ontology should be seen. In addition to this problematic question, there is an additional question, namely: how should reduction be actually pursued? There are different views about the nature of inter-theoretic reductions, but what is striking is that a clear suggestion about how any reduction could be achieved in the case of the mental and the physical is lacking. For a programmatic proposal of such an account, one can consider the works of Bickle,\textsuperscript{1149} Paul Churchland,\textsuperscript{1150} or Patricia Churchland.\textsuperscript{1151} We could be interested

\textsuperscript{1146} This is Ramberg’s (1999) insightful interpretation of Davidson’s general intention.
\textsuperscript{1147} Bickle, 1998, 29.
\textsuperscript{1148} Bickle, 1992, 217, my emphasis.
\textsuperscript{1150} Churchland, 1995.
\textsuperscript{1151} Churchland, 1986.
in the details of these proposals, but we could also take Paul Churchland’s word when he asks: “Can we reconstruct all known mental phenomena in neurodynamical terms?” and answers: “Not at the moment, we can’t. Not by a long shot.” In a work which attempts to defend the irreducibility of the mental, it suffices to note that it is absolutely clear that currently the best suggestions do not come even close to establishing mental–physical identities. No matter how the successful reduction is actually carried through, it should be such (according to Paul Churchland) that it reconstructs, in neurodynamical terms, all of the mental phenomena antecedently known to us, and it should teach us some things about the nature and behavior of mental phenomena that we did not already know.

In considering whether Davidson’s arguments can challenge a version of new-wave reductionism, let us briefly evaluate Bickle’s model of reduction, which is described in his Philosophy and Neuroscience: A Ruthlessly Reductive Account. This is a good proposal to consider, because Bickle is a philosopher who also has an interesting scientific agenda. Such a position captures well the confusion concerning the nature of philosophy and its relation to science, a confusion which in my opinion is an extremely relevant issue for modern philosophy of mind and its future. Bickle is one of the most devoted of contemporary reductionists. In 1998 he suggested a model of new-wave reductionism which was meant to be an improvement to the Hooker-Churchland model of reduction. In 2003, Bickle described yet another development in his theory, the key term now being new-wave metascience, on which the reductive model is based. The main point of this approach is that we should listen only to scientists when evaluating the results and significance of our work. New-wave metascience: “eschews all traditional concern with ontology and ‘metaphysics’…” and its “job is simply to illuminate concepts like reduction as these imbue actual scientific practice.” This raises the question of whether Bickle is on the same page with other philosophers. He is not. As reasonable as the approach of new-wave metascience may sound, it is absolutely clear that the guiding intuition behind it differs radically from a view that does not see philosophy and science as continuous, but rather as distinct enterprises. Bickle’s position is straightforward:

---

1152 Churchland, 1995, 211.
1154 Bickle, 2003, 32. What does the expression “illuminate concepts” mean here? Given how empirically-orientated Bickle’s attitude is, surely he is not doing conceptual analysis!
New wave metascience doesn’t treat traditional ontological questions as ‘cognitively meaningless’, just pointless to pursue as serious intellectual work because they amount to nothing more than disagreements over bare intuitions, with no agreed-upon evidence for resolving them. But counts as “serious intellectual work”? In philosophy one rarely offers evidence but instead arguments to defend one’s views. It is obvious that many traditional ontological questions can be dismissed if Bickle’s reasoning is followed, but why should philosophers agree with the ideology of new wave metascience? Von Wright, for example, thinks that the task of a philosopher is to clarify conceptual intuitions and evaluate critically the intellectual currents of her times. It is precisely our intuitions, so despised by Bickle, that provide our starting point and to which we must refer in the end. Shouldn’t a philosopher discuss problems the possible solutions for which are not based on agreed-upon evidence? Not according to Bickle:

Scientific philosophy as it will be practiced in this book… is heir to a brilliant tradition spanning most of the 19th and 20th centuries, until the end of the latter, when ‘speculative metaphysics’ infected even philosophy of science and philosophy of particular sciences with ‘external’, ‘pragmatically fruitless’ debates that turn on nothing more than clashing intuitions. Bickle praises scientific philosophy and his proposal is meant for scientifically-inspired philosophers. The meanings of these definitions are not explicated. Bickle nevertheless claims that scientifically oriented philosophers should dismiss certain questions because they are “meaningless”. This view, as Bickle acknowledges, is heavily influenced by Carnap and his distinction between external and internal questions. I believe we should notice how strongly Bickle’s background intuitions guide his reasoning and philosophical work. Here the conflict between science and philosophy, emphasized in chapter one, is clearly visible. We are told that philosophers should trust neuroscientists, who understand neuroscience much better than philosophers, in order to evaluate what their work has accounted for. But if we take this line, precisely who should we trust? Bickle is ready to reduce “minds to molecules” but, just to take one example, Benjamin Libet, a very scientifically oriented thinker and an acknowledged pioneer in the study of consciousness claims that:

As a neuroscientist investigating these issues for more than thirty years, I can say that… subjective phenomena are not predictable by knowledge of neuronal function…. conscious mental phenomena are not reducible or to explicable by knowledge of nerve cell activities. You could look into the brain and see nerve cell interconnections and neural messages popping about in immense profusion. But you would not observe any conscious mental subjective phenomena.

---

1155 Bickle, 2003, 35.
1156 Bickle, 2003, 39.
1157 Libet, 2004, 5. We should recall that Sir John Eccles, the Nobel Prize winning neuroscientist, was a dualist!
This is what also von Wright claims. I believe we can safely expect that for each scientist agreeing with Bickle, we can find a scientist who disagrees with him. Partly due to sociological reasons, there is no consensus among “scientifically minded” philosophers about the problems of mind. Who are we to trust according to the new wave metascience? One could answer that the results of experiments speak for themselves, but isn’t the interpretation of experiments strongly guided by “intuitions”, which Bickle so gladly ridicules? This, the clashing of intuitions based on a set of different convictions and different intellectual agendas, is precisely what the debate between Bennett, Hacker, Dennett and Searle is partly about. How the results of neuroscientific research should be interpreted is a question to which philosophers can and should make a contribution.

In Bickle’s case it has to be questioned whether there is even a possibility for a philosopher to evaluate the correctness of Bickle’s account. The prospects are not good given that Bickle warns: “If the scope of my concern is too narrow for your philosophical sentiments, so be it. Scientists don’t give a hoot for philosophers’ Very Strong Modal Intuitions about kind identity across possible worlds, and their enterprises are doing just fine.”\footnote{Bickle, 2003, 134.} He continues by noting: “It might be fashionable in post-Kripke philosophy to insist that all identity claims hold necessarily, but fortunately scientists don’t bother reading Kripke and keep right on making and testing identity claims that purport to hold in the real world.”\footnote{Ibid.} I think we should find these views quite incredible. They are narrow-minded and show that, as suggested in the first chapter, certain naturalists simply do not care anymore about philosophical arguments and do not think that they could have any relevance for the contemporary mind–body problem. It seems to me, as this specific view of Bickle shows, that Bennett and Hacker are precisely right in their criticism that scientific philosophy is seriously threatened by the possibility of conceptual confusion. They claim that views of neuroscientists are filled with confusion and incoherence.

It would be quite naïve to think that current science would be immune to such problems. It is symptomatic that some reductionists simply ignore the criticisms of philosophers that arise from a consideration of conceptual aspects. By dismissing the relevance of philosophical arguments, Bickle seems to think that new-wave metascience simply is free of conceptual problems. Dogmatically, it is only scientists who are allowed to evaluate the relevance and correctness of scientific arguments. As far as I can see, this attitude is a perfect example of the scientististic attitude of our times, and it seriously belittles the
importance of philosophy as a human activity. This attitude was something of which Wittgenstein in his times warned about, and of which philosophers like Bennett, Hacker and von Wright have reminded us recently. Bickle’s attitude testifies to the fact that these warnings and reminders have not been empty. Given Bickle’s view, it should perhaps be worthwhile to consider whether he should be taken seriously at all from the perspective of philosophy. If he “does not give a hoot” about the philosophical arguments that could be raised against his position, why bother presenting them at all?

Leaving the ideological aspects of new wave metascience aside for a moment, it should be asked how plausible its reductive project seems to be. The purpose is to reduce “minds to molecules”. It seems to me that for this to happen, it is required that psycho-neural reductionists actually provide psycho-neural reductions. The purpose of new wave metascience is to avoid all reference to psychological entities and achieve the explanatory scope of current psychology by their “simpler” means. The goal is to have a purely physical account of behavioral causes. New wave psycho-neural reductionism is defined as “the prediction that as mature theories develop in psychology (T_Rs) and neuroscience (T_Bs), images (I_Bs) of the former will be constructible within the models of the latter.” This suggestion tells us something about the expected relation between the mental and the physical, but what it tells is not much. Here we have basically a suggestion that an unspecified super theory of the future will be such that an unspecified psychological theory can be constructible in it. But why believe that such a neuroscientific super theory is possible? Bickle’s answer is that if we consider the data so far we can see, especially if we keep our philosophical arguments to ourselves, that reductions are already being done and there are more to come. This of course is not an argument for a super theory, but at best a reason to have faith in the possibility of such a theory. It is a dogmatic belief in scientific progress and not an argument.

In his book Bickle describes the reductions that neuroscientists have already achieved. The main case which he presents is the description of how memory consolidation in rats has been reduced to the molecular mechanisms of long-term potentiation (LPT). I am not going to discuss the details of this because, needless to say, my knowledge and understanding

---

1160 I assume that Bickle is not very interested in the conceptual distinctions between reasons and causes, behavior and actions. I base this assumption on his views concerning the irrelevance of philosophical analysis.


1162 Another example: According to Bickle a successful reduction of actual psychology to neurobiology is a description of the cellular mechanisms of short-term and long-term learning as revealed by studies of the sea slug, *Aplysia Californica*. We will consider the relevance of this reduction below. For other examples of suggestions of successful reductions see Bickle, 2008.
of molecular neuroscience is not sophisticated enough and the details can be found in Bickle’s book. I admit that I am not the kind of “scientifically-inspired” philosopher for whom Bickle’s book is meant. But neither are von Wright or Davidson. I believe we can recognize the general problems of new-wave reductionism without taking any stance on the neuroscientific details provided by the position. We have no reason to doubt that these details might not be correct. It could be complained that here we have a typical example of a philosopher not understanding and not bothering to study the results of neuroscience, thereby missing the important reductive results. This claim is unjustified because, while ignoring the details, we can grant what Bickle says and, as we will see, still raise criticisms about his conclusions.

Let us grant what Bickle says, namely that there is no currently active psychology of memory consolidation, and that this is because the link between memory consolidation and molecular mechanisms of LPT has been found. I would like to propose four problems which are in need of solution if new wave psycho-neural reductionism is to be a relevant contribution to the question of whether our mental vocabulary is autonomous and irreducible. On the one hand, these problems query whether new wave reductionism is a breakthrough when it comes to the question of mental–physical reducibility. On the other hand, they show how Davidson’s arguments in favor of irreducibility are, pace Bickle, also still relevant with respect to a recent reductive account. The only way to find them to be irrelevant is to ignore them, as Bickle indeed does.

Whereas Davidson’s arguments are considered in Bickle’s earlier writings, new-wave metascience is a position which does not need to care for these kinds of arguments or for any philosophical arguments whatsoever. But as a challenge for the new-wave model, certain questions can be raised: 1) What do the results of the molecular mechanism of LPT in rats or the cellular mechanisms of the sea slug tell about the prospects of reducing the totality of human psychology to neuroscience? 2) Would an “ontological reduction” of mental phenomena show that our mental vocabulary is dispensable? 3) Are there reasons to draw eliminativist conclusions from the claims of new-wave reductionism? 4) Does Bickle’s model show that Davidson’s thesis about the anomalism of the mental is false? Let us consider these

---

1163 Although I have to note that in a personal discussion, Joe Levine, who in his Purple Haze has famously argued that the problem consciousness may not receive a physicalistic answer, claimed that nothing said by Bickle is true and that Bickle is “lying”. This was interesting charge from Levine’s point of view, but the motivation for it was understandable. Levine, for reasons unclear to me, thinks that psychology is autonomous and irreducible to neuroscience and therefore he of course cannot accept Bickle’s views. As a sideline in the “sociology of philosophy”, this can be to be found interesting.

questions in turn. 1) As long as non-human animal cognition is thought of in purely mechanistic terms it is not surprising that there exists an explanation of it in terms of molecular mechanisms. Davidson agrees that some cognitive functions can be correlated with physical mechanisms. His claim is that propositional attitudes cannot be correlated with specific brain-states. The results described by Bickle are not a challenge to this view and they do not hold any promise that propositional attitudes would be reducible in a similar way as memory consolidation is reducible to LPT. I think we should also consider whether the learning process of a sea slug counts as a psychological phenomenon. Bickle does not suggest at all how a complete psychological theory could be constructed, what it would look like, and how it would be reduced to a neuroscientific theory. But these questions require answers if Bickle is attempting to come up with a neuroscientific theory which would achieve the explanatory scope of psychology. The hope that a future neuroscientific theory will replace existing psychological theory is empty if we are not told how to construct an image of P (psychological theory) in the conceptual framework of N (neuroscientific theory). As Bickle notes, the relation between the original psychological theory and its deduced image is something which neither Hooker nor Churchland has ever provided. How this construction of an image should be done for the whole psychological theory remains a mystery, and the few existing examples of successful reductions, like the sea slug case, are not even a part of the answer.

2) Given Davidson’s views, it should be obvious that he could accept, for example, the claim, that memory consolidation is LPT and, ultimately, identical to the molecular mechanisms of the latter. Something like this seems to be implicitly suggested in Davidson’s “Material Mind”. From an ontological point of view they are the same phenomena, described at different levels. Does a conceptual reduction follow from this? Bickle argues that ontological and conceptual reductions must be closely related. Ontological conclusions follow from the reduction relation obtaining in a given case. In this sense ontology is secondary with respect to reductive results, although it is closely tied to them. What does this mean with respect to our common-sense understanding of mental phenomena? Are there such things as beliefs? There are, if the vocabulary which uses them reduces to a future science. But if this happens, it seems that our ordinary concepts become otiose. Is there then any possibility for an independent and autonomous existence for our mental vocabulary? It seems not; we are justified in using this vocabulary but it is a candidate for reduction, and the claim is that when

---

a reduction is accomplished, the original vocabulary becomes useless. In this sense the ontological and conceptual levels are closely connected, and both are dependent on the possibility of their reduction to some future neuroscience. When we show that memory consolidation is LPT, we lose the need to use the former term in explanations, and in the ideal yet to be imagined case, a whole vocabulary of higher level terms becomes redundant with respect to the vocabulary of molecular level.

On the one hand it is claimed that through a reduction of psychology we should achieve the same explanatory scope that psychology has now, and that the explanatory framework would have been achieved by simpler means. This is the goal of scientific reduction. On the other hand, once the mental has been reduced, we see that psychological explanations pale in comparison with neurological explanations. It would be ill-advised to keep on using these inferior explanations. But I think we should consider whether a neuroscientific explanation is really simpler than a folk-psychological explanation. Was Ryle wrong when he claimed that “We know quite well what caused the farmer to return from the market with his pigs unsold. He found that the prices were lower than he had expected. We know quite well why John Doe scowled and slammed the door. He had been insulted.”

Are neuroscientific explanations simpler than these? It is difficult to see how a detailed neuroscientific description of what happens when John Doe scowls and slams the door would be a simpler explanation than the conclusion that Doe was insulted.

What about the claim that reductions from minds to molecules make psychological explanations useless because these explanations are inferior when compared to their neurological reducers? This is a strange claim for three reasons: A) If psychological explanations are different kinds of explanations than causal explanations, how could the latter make the former dispensable? B) Let us suppose that token-identities could be established. What would be consequence of this for explanations? Without laws describing the relations between minds and molecules, the psychological level, with its respective generalizations, remains explanatory in itself. What reasons have we to believe that nomological connections exist between these levels? C) When the deficiency of psychological explanations is being pronounced, we should consider how the term “explanation” is understood.

As far as A) is concerned, Bickle is surprisingly insensitive to the possibility that rational explanations in terms of our mental vocabulary could be different kinds of explanations than the causal explanations of neuroscience. The alleged difference between

---

rational and causal explanations can be summarized by considering two theses from Kim. A reference to Kim in this context is instructive because it shows that a reductionist can also appreciate the difference between two kinds of explanations. Kim describes a mode of understanding which is not causal-predictive in six theses, two of which are:  

Thesis 1) Each of us has a need to understand, make sense of, our own actions… it is an essential part of our nature as reflective agents that we need to render our own actions and decisions intelligible to ourselves.

Thesis 4) The purpose of this mode of understanding is not predictive; nor is it aimed at acquiring knowledge of the causal mechanisms leading from our desires and beliefs to actions.

Kim made these points long before he started to defend a more reductionist position. More recently, perspicuous comments on the differences between rational and causal explanations have been made by Stoutland. He notes, in a similar way to Kim, that: “The main function of rational explanation is to enable us to make sense of the people we encounter so that we can respond to each other as how we are, how we feel, and the ways our lives are shaped.”

This is in essential respects the point of Kim’s first thesis. Stoutland agrees also with Kim’s fourth thesis: “A rational explanation… does not aspire to specify any underlying mechanisms… that explain the agent’s behavior, which means that a philosophical account of rational explanation will not attempt to specify any explanatory mechanisms.” I believe we could say that the purpose of rational explanations is also predictive, but that the predictive power of these explanations does not depend on the fact that they increase our knowledge of the causal mechanism(s) behind the explanation. Rational explanations predict in a very different way than causal explanations. Bickle seems to admit that he is interested only in one kind of explanation, namely the causal one. But this does not tell us whether he thinks that all explanations should be understood as being causally mechanistic. If the answer is affirmative, then it is not surprising that causal mechanistic explanations in psychology could be reduced to causal mechanistic explanations of the molecular level. Bickle writes: “[…] contexts determine whether a causal-mechanistic explanation, given the current state of

1168 See Kim, 1984, 316, 318.
1169 Stoutland, 2005, 145.
1170 Ibid.
1171 See, for example, Bickle 2003, 114.
scientific knowledge, is *genuinely causally mechanistic*. In scientific contexts, typically only one is treated as such at any given time, namely, the one genuinely available at the lowest level….”

*If* psychological explanations are causally mechanistic, then following Davidson, we can argue that they belong to the same conceptual domain with the physical and there is no principled obstacle to reduction. But if the nature of psychological explanations is other than causal-mechanistic, then it is questionable whether their explanatory power could be reduced to causal mechanistic explanations at the lower level. It seems that rationalizing explanations proceed and can succeed independently of any knowledge about the internal mechanisms of the agent whose actions are being explained. Rationalizing explanations may not be extensible to mechanistic explanations, but this does not affect Davidson’s argument because his claim is that *they are not supposed to be so extensible*. Your knowledge of someone’s mental properties does not tell you anything about the internal mechanisms involved; nor is this kind of knowledge required in order to explain something in terms of mental properties.

A brief hint suggests that Bickle is willing to accept also other forms of explanation than just the causal mechanistic variety. He notes: “There are… numerous contexts, both everyday and scientific, where the explanatory power of the cellular / molecular neuroscientific causal account isn’t important…. In everyday contexts… we are more lenient about the number and variety of explanations we offer and accept.” But it is unclear what, according to Bickle, the status of these “everyday explanations” is. It is likely that they should be counted as second grade explanations because they are not “scientific” explanations. It is clear, though, that at least in the case of psychological explanations, their *explanatory power* is less than that of their neuroscientific challengers. We are told:

> […] when we fix our gaze on aspects of scientific practice… we see that psychological explanations *lose their initial status as causally-mechanistically explanatory*… within scientific practice, psychological explanations *become otiose* when the type of cellular / molecular explanation… is achieved.

Bickle claims that neurobiological explanations make psychological explanations pointless. The correctness of this claim could be granted if psychological explanations were seen as causal-mechanistic, as Bickle does. Unfortunately, Bickle does not consider the possibility that psychological explanations are not causal mechanistic, and therefore he does not really

---

1172 Ibid.
1173 Bickle, 2003, 110.
consider the kinds of arguments that Davidson and von Wright have used to defend the autonomy of rational explanations. Bickle acknowledges that psychological concepts seem to be “intractably dualistic”. What he does not recognize, or at least does not discuss, is that while this feature of these concepts may be explanatorily relevant, there may be essentially different kinds of explanations as Davidson, von Wright and Stoutland suggest. Given Bickle’s admittance that “everyday” practices differ from scientific ones, it remains unclear what it means that certain explanations are otiose or pale in comparison to some other explanations. In the case of rational explanations, do we have a single example showing that the explanatory power of this kind of explanation becomes otiose because a neuroscientific explanation has replaced it? Do we have a single example showing that the explanatory power of rational explanation pales when compared to a neuroscientific explanation which is meant to explain the same thing?

As far as question B) is concerned, there is no reason to think that new wave reductionism could provide mental–physical identities which would be secured through laws. This possibility is in fact denied, and therefore new wave reductionism differs from a mind–brain identity theory. If mental–physical identities do not exist, how could neuroscience explain the same events in the same way as psychology does? The generalizations of the latter vocabulary would be absent in the former vocabulary. Moreover, and this is Davidson’s question, without laws what would lead us to believe that our observations of the connections between the mental and the physical tell something about the future cases? As he notes: “what we want as a law of correspondence is not an accidental, but one which we have reason to believe, will apply to cases which we have not yet examined.”

What about question C)? Let us suppose that in reply to the question “Why is Georg calling his brother?”, the following answer is given: “He hasn’t heard from his brother for two weeks and he is worried because his brother is an alcoholic.” This is certainly a reasonable explanation, and in most cases we would be satisfied with the answer; we need to know nothing more to be epistemically satisfied. The explanation makes sense of the situation. Kim has noted: “When we look for an explanation of an event, we are typically in a state of puzzlement, a kind of epistemic predicament.” Is the epistemic puzzlement in those cases when we are looking for an answer in psychological terms such that it would decrease if we were given an answer in neurobiological terms? I think this seems very unlikely. Perhaps our overall puzzlement would in fact increase, because we could start to wonder what an answer

---

1175 Davidson, 1964, 47.
1176 Kim, 1989a, 254.
in terms of molecules has to do with Georg and the phone. Kim claims also: “We expect our reductions to yield simpler systems – a simpler system of concepts, or simpler system of assumptions, or simpler system of entities.”\textsuperscript{1177} In what sense would a reduction which replaces the psychological explanation with a neuroscientific explanation yield a simpler explanatory system? One of Davidson’s claims is that our common-sense understanding of ourselves is so valuable to us precisely because it is, in a sense, a very simple mode of explanation. It is highly unlikely that a molecular explanation would be something that we could ever use to describe and explain actual everyday situations which occur outside the artificial environment of a laboratory. We \textit{already know} why John Doe scowled and slammed the door. Are we, in this case, \textit{in need of a further explanation}? Are we puzzled? There is something implausible in the suggestion that we would need a deeper explanation of why John Doe slammed the door than the one we already have. What would this “deeper” explanation tell us of the situation in which we encounter John Doe?

How we think about the status of common sense explanations depends partly on our understanding of the general nature of explanations. If, together with Davidson, we think that explanations are always interest relative, then we can hold on to the idea that \textit{different kinds of explanations} can be given of the same event – or of phenomena – and that these explanations are independent of each other. We have different kinds of epistemic expectations, and the explanations we are ready to accept depend on the expectations we have on each occasion. The explanation that Georg’s brother is an alcoholic may be just the kind of explanation that we were looking for. Common sense mental explanations usually take place in real situations, where the complex context may give a clue as to why certain kind of explanation is better than some other in the specific situation in question. If we are always looking for a \textit{causal-mechanistic explanation}, we surely are looking for \textit{a certain kind of answer} to the question “Why is Georg calling his brother?” and we are seeing the whole situation in a certain kind of way. But would we be satisfied, in a real situation, if our question were answered in terms of molecular biology? Bickle notes:

\begin{quote}
The explanation, ‘Kurt remembered that telephone number today that I relayed to him yesterday because he rehearsed it mentally fifteen times without retrograde interference for thirty minutes after he heard it’ pales in comparison to one that appeals to activity-dependent molecular (including molecular genetics) mechanisms occurring in millions of selective neurons.\ldots \textsuperscript{1178}
\end{quote}

\textsuperscript{1177} Kim, 1998, 96.
\textsuperscript{1178} Bickle, 2003, 114.
Bickle is here talking about the different ways in which we explain or could explain the functioning of the memory. I think it is not clear in this case whether the former explanation “pales in comparison” to the latter, since it is not clear what it means to say how explanations should be compared. What is clear is that if our epistemic interests are geared toward a kind of answer like “Georg’s brother is an alcoholic”, we are not satisfied with an answer which refers to molecular facts – and in this case a comparison between these answers would be utterly inappropriate. Why believe that a mechanistic explanation, even if it could be had, would satisfy the expectation that the questioner has on this occasion.

Bickle’s views about the nature of explanation are not clear. Should we always aim for a causally mechanistic explanation at the lowest level, are explanations interest relative, and do rational explanations differ essentially from causal mechanistic explanations? These important questions remain without an answer. Bickle claims that: “Scientists tend to do just fine with a rough-and-ready understanding of what counts as an ‘explanation’ and what distinguishes a ‘good’ one from a ‘poor’ one.”¹¹⁷⁹ This being the case, scientists do not need to offer an account of explanation and they do not need to further analyze the concept of “explanation”. But this view seems to suggest that scientists get to decide what counts as a good explanation, because they already have a “rough and ready” understanding of what explanation is. It is very curious that a “rough and ready” everyday explanations of mental phenomena are treated as second-grade. To suggest that scientists get to decide what counts as an acceptable explanation is merely to repeat the ideology of new-wave metascience. Perhaps the claim is true in those cases when a causal mechanism can be explained in terms of a lower level causal mechanism. In these cases, scientists may agree that the lower level explanation is “better.” As Bickle argues: “In scientific contexts, typically only one [explanation] is treated as [genuinely causally mechanistic]... namely the one genuinely available at the lowest level.”¹¹⁸⁰ But why think that scientists get to decide whether or not an explanation which refers to the character of Georg’s brother counts as a good answer to question “Why is Georg using the phone?” Bickle claims that it is utterly wrongheaded to assume that psychological explanations would remain causally explanatory if lower level neurobiological explanations explained the same behavioral data. It is unfortunate that Bickle does not discuss the positions of von Wright, Stoutland and others who argue that it is a conceptual confusion to claim that neurobiological explanations could explain human actions.

¹¹⁷⁹ Bickle, 2003, 57.
To claim, like Bickle does, that we need not have a detailed philosophical view about explanation in order to see that psychological explanations become pointless when the neurobiological explanations are in place sounds curious. Once the lower level explanations are found, psychological explanations become “explanatorily impotent and empty”. We are told: “The cellular / molecular neurobiological account explains many key causal processes that the psychological account is either completely blind to or leaves as input-output black boxes. In other words, it explains events that the psychological account leaves unexplained.”1181 Bickle does not say a word about a view like Kim’s, and more obviously those of von Wright, Davidson and Stoutland, who claim that the purpose of the psychological mode of understanding is not predictive; nor is it aimed at acquiring knowledge of the causal mechanism or of the “key processes” that Bickle refers to. To say that neuroscience explains events that psychology leaves unexplained is not a breakthrough, since the same can be said of psychology with respect to neuroscience.1182 Which explanatory scheme is then generally simpler or has a wider explanatory scope? I suggest that this question does not make sense and should therefore not be asked.

3) The answer to the question of whether new wave reductionism leads to eliminativist results is straightforward: it does not. Pace Bickle, it is not the case that neurobiological explanations render psychological explanations “pointless”. The main reason for this is the fact that new wave reductionism does not appreciate the distinction between rational and causal explanations. New wave reductionism is also silent about the conceptual distinctions between behavior and action and between reasons and causes. We have seen no good reasons to believe that the explanatory power of our mental vocabulary would become otiose or that it would pale when compared to the power of neurobiological explanations. In fact, Bickle

1182 Cf. von Wright (1998, 37): “If the correspondence [between reasons and neural processes] is well established, the neural state in question may be regarded as a reliable sign or symptom of the existence of the reason. As long as the correlation remains a scientific hypothesis, the neural state fulfils this role of a symptom. Only in the very unlikely case that the hypothesis became so well confirmed that we would be extremely reluctant to drop it when faced with seemingly contrary evidence could we conceivably use the neural state as a criterion of the agent’s having a certain reason for action. And even then the criterion would only be one among many, and its usefulness in attributing to agents reasons for their actions would depend upon how well it contributed to our understanding of the agent as a person and to the agent’s understanding of himself.” It certainly seems to be the case that psychological explanations have a dimension which matters for the subject himself. An agent can understand how he became jealous because of the events which took place and which he understood in such and such way. Somebody may understand why he was offended because he, for example, misunderstood something. Both these explanations would be expressed in terms of propositional contents. These could be paradigmatic examples of events which can be explained in our mental vocabulary but which the neuroscientific explanations leave unexplained. How an agent could understand himself, as a person, in terms of molecules?
admits that current science does not eschew all mental or psychological causal explanations. Moreover:

[...] psychological causal explanations still play important heuristic roles in generating and testing neurobiological hypotheses...they suggest where to look for the key cellular and molecular mechanisms and how best to construct behavioral paradigms that will generate useful tests of hypothesized cellular and molecular mechanisms.\footnote{Bickle, 2003, 114.}

The general lesson that Bickle wants to teach philosophers of mind is that they should “re-orient their interests down levels in the neurosciences.”\footnote{Bickle, 2003, xiii.} This view is the opposite of von Wright’s, who argues that the mental has an epistemic priority in relation to the neural. The role that Bickle grants to psychology is a secondary role; psychology and our mental vocabulary are useful for testing neurobiological hypotheses. But if psychological explanations are useful in generating and testing neurobiological hypotheses, doesn’t this mean that, from an epistemological point of view, they indeed enjoy a privilege over the hypotheses of neurobiology just as von Wright claims?

Let us suppose that we have framed a neurobiological hypothesis, for example, about seeing and we attempt to create a visual sensation by manipulating the molecular level in accordance with the hypothesis. How could we know that the test-subject actually sees something if we stay at the molecular level? Or to use another actual example: sometimes it happens that completely paralyzed people are thought to be in comas. In some cases, a brain scan may show that their brains are functioning normally. After the scan is done, a person who was thought to be in coma is given a technique to communicate with the outer world. What, then, is the final verification method – is it the brain scan or is it the actual communication produced by the patient? I think there is still a long way to go before we would be willing to accept that a brain scan would convince us of the presence of consciousness if all other signs were absent. Perhaps this will change in the future, which would mean that our notion of consciousness had changed; the state of the brain would become the ultimate criterion of the presence of consciousness. On the other hand, perhaps we will never be ready to accept the presence of consciousness without behavioral manifestations. An interesting way in which states of brain could become a criterion for consciousness could be the following. An assumed coma patient would be given simple instructions, for example, to think about the performance of a certain task. If this thinking could be linked to a specific pattern of brain activity, then the patient could be instructed to
think about the performance of a task with the intention of producing “yes” and “no” answers. A particular pattern of brain activity would then be interpreted as an affirmative answer, and as a consequence simple questions could be asked from the patient. Recent research actually indicates that fMRI could be used as a form of communication. But would the communicating state of the brain then count as a brain state or as behavior? If, as von Wright claimed, the result of sheer willing cannot be action, how should we understand the “fMRI-answer” to a certain question? In normal conditions, “answering” counts as an action. If the answer could be produced by sheer wanting or willing, would this invalidate the answering as action?

Can we conceive of a situation where the heuristic role of psychological explanations could be completely dismissed? If we agree with von Wright, as I think we should, that the neural cannot be a criterion for the mental, it seems that on a very general level, which is also at the same time a very fundamental level, the psychological keeps its autonomous status and cannot be eliminated. As von Wright has claimed, we cannot be sure whether neurological phenomena are relevant for the mental phenomena for which we take them to be relevant if we have no independent way to check whether the mental phenomenon in question occurs. The status which von Wright grants to the mental is indeed a fundamental one; a neural hypothesis could not be verified at all in the absence of the relevant mental phenomena. This I take to be a philosophical observation of the utmost importance, and I cannot see how it could be rejected. Bickle grants a small role for psychology, but his reason for doing this is that the “real” causal mechanistic low-level explanations have not yet been discovered. He claims that psychological explanations become useless once the “real” explanations are found. But this assumption collapses if the main purpose of psychological explanations is normative evaluation, and not causal explanation and prediction. Bickle does not discuss this possibility.

We have seen how Davidson’s arguments could be used to refute a type-identity theory, and Bickle admits this possibility. A type-theory could lead to eliminativism if all mental phenomena could be correlated with physical phenomena; we could possibly lose interest in explanations in terms of mental concepts. Bickle acknowledges the importance of Davidson’s arguments against Nagelian reduction, but denies the importance and relevance of these arguments to the new-wave model, because in this model bridge-laws are not required. But as I have emphasized, Davidson’s views about the irreducibility of the mental are by no

1185 See, Owen, Schiff and Laureys, 2009.
1186 For an interesting discussion of how Wittgenstein’s insights about the connection between mental states and behavior could be relevant in the cases where patients were in persistent vegetative states, see Gillett, 2001.
means exhausted by his claim that there are no strict laws between the mental and the physical, and the arguments for irreducibility may be effective against new-wave model.

What kind of reductive view is the new-wave model, if it does not require strict connections between the mental and the physical? How a new wave reduction between the mental and the physical might actually happen remains unanswered. We are not told how an “image” of psychological theory would or could be constructed in the vocabulary of neuroscience. What Bickle provides is only the prediction that this is going to happen. I think that even if new wave reductionism avoids nomological connections between two domains and even if a neuroscientific image of psychology could be constructed, there is still a major part of Davidsonian irreducibility to be answered – namely the claim that our common-sense understanding is not an optional part of our conceptual resources. This is a Davidsonian challenge that Bickle does not evaluate at all. He tries to defuse Davidson’s strict-law argument in his earlier works and, apparently thinking that this is all there is to Davidsonian irreducibility, he does not mention Davidson’s criticism in his later work where the model of new wave reductionism is further developed.1187 When arguing against the irreducibility of the mental, Bickle focuses on arguing against functionalism, but the functionalist arguments he refutes are not the same kind of arguments as those presented by Davidson, and the charge against functionalism does not really work against Davidson’s position. So, unfortunately, Bickle leaves entirely without discussion an important part of Davidson’s criticism against the possibility of reduction. It is not certain that Bickle himself recognizes this because he notes: “[the Davidsonian] challenge boils down to the charge that intentional psychology and the physical sciences are composed of very different sorts of linguistic items: nonlaws versus laws. It is this difference that is supposed to scuttle the possibility of reduction.”1188 But this is not all that there is to irreducibility.

It is telling that Bickle does not say anything about the alleged normativity of the mental, because this is perhaps the feature of the mental which is the most difficult part for a neuroscientific account to solve. By not discussing it, Bickle of course avoids the problem. The same applies, to some extent, to Patricia Churchland who writes: “Davidson… has proposed antireductionist arguments, and though his work has been widely discussed in the philosophical literature, I find Dennett… Fodor… and Pylyshyn clearer and deeper.”1189 We cannot of course blame Churchland for thinking that Davidson’s arguments are not clear or

1188 Bickle, 1992, 228.
1189 Churchland, 1986, 379.
deep; opinions about these matters differ. What should be noted though is that it is quite easy to dismiss any arguments by declaring that they are not clear enough or deep enough, and therefore the strategy adopted by Churchland is perhaps not the best one for argumentative purposes. Be that as it may, it is clear that Dennett’s, Fodor’s or Pylyshyn’s arguments are not based on the same views as those of Davidson, and I think that therefore one cannot leave out Davidsonian arguments on the grounds that one has considered the others.

Bickle’s example of the reduction of memory consolidation to molecular processes is well chosen because, or so I think we can conclude, it is not a rational process and therefore the problem of normativity does not arise. I conclude that since Bickle’s model of reduction remains silent with respect to the problem of the normativity of the mental, it is not secured against Davidson’s arguments. We can note here that the same applies to Kim’s “functional model of reduction”, which is another recent attempt to reduce the mental to the physical.

4) Does new wave reductionism show that the thesis of the anomalism of the mental is false? It does not. Why is this important? Davidson’s claim in “Mental Events” was that there are no strict deterministic laws on the basis of which mental events could be predicted. I have argued that the suggestion is more modest than most critics have taken it to be, but it is also very ambitious because it marks a principled difference between mental and physical events. If the suggestion is true, a principled autonomy of mental concepts is secured. Nomological independence removes the mental from the reach of the nomological net of a physical theory.

---

1190 The fact that the examples concern non-human animals makes this even clearer. When discussing the “psychology” of sea slugs, we do not really need to worry about the problems of rationality, at least not about the kinds of problems that we encounter when discussing the nature of human rationality.

1191 What this concretely means will be discussed in the next three sections. It should be noted that there is a brief suggestion which could be taken to indicate that Bickle has considered the normativity argument. He says (2001, 252): “‘Anomalousness’ in the sense Davidson stresses for mental predicates does not block a potential new wave reduction of theories because of the nonlinguistic (model theoretic) formulation of theory structure and relations.” As far as I understand this view, it is meant to suggest that since scientific theories should not be understood as “sets of sentences” but rather as model theoretic structures, the remarks about the irreducibility of psychological concepts is not relevant to the question of whether psychology, as a theory, reduces to neuroscience. This view raises a host of problems. First, the question of whether common sense psychology can be understood as being a theory is problematic. Second, accepting the view requires generally accepting a structuralist view about scientific theories. The correctness of this suggestion is a question which cannot be dealt here in any detail; nor would I have competence to consider it. Third, we are required to accept the view, put forward by Paul Churchland, that knowledge representation and exploitation is not essentially linguistic. However, these suggestions are, as Bickle (1992, 228) himself notes, “extremely programmatic”. Fourth, suppose that we grant all of this. This would be very generous since it requires a whole new way of thinking about reductions as well as scientific theories. But even if we accept all this, there remains the question about the current explanatory power of psychological concepts. We understand ourselves and others in terms of mental concepts because of the conceptual connections between these concepts. The connection between reasons and actions exists at the level of propositional contents. This explanatory power would collapse if we accepted the view that knowledge is not represented and manipulated in terms of linguistic items.

1192 See Kim, 1998. As we will see in section 3.3.3.1, Kim is better aware of the Davidsonian criticism than Bickle is; yet this does not show that his account can defuse the Davidsonian challenge.
Instead of trying to prove that the anomalism of the mental is true, we could ask for reasons to think why it might be false; the thesis of anomalism seems to fit with our common-sense intuitions. New-wave reductionism does not offer such reasons, and whatever the fate of new-wave reductions turn out to be one could still argue that, because of the anomalism of the mental, the mental remains unpredictable. This is one way to formulate the autonomy of the mental; mental reality remains unpredictable and this is not the case with physical reality.

In conclusion, the following can be said about the new wave challenge for Davidson’s position. If we are trying to decide whether the mental lives of humans can be reduced to brain-processes, new-wave psycho-neural reductionism is, despite the fierce advertising, not a significant breakthrough. First, the examples of how “psychological phenomena” in sea slugs have been reduced to the molecular level do not tell us much, if anything, about the prospects of reducing the complex cognitive states of humans to lower levels of physical reality. Second, the prediction that a future super-science will reduce psychology is based on mere faith concerning the possibilities of science and scientific progress. Third, new-wave psycho-neural reductionists remain silent about the nature of explanation. This raises many questions: what is the relationship between rational and causal explanations? How can explanations replace each other? How to evaluate the simplicity of explanations? To state, as Bickle does, that scientists have a “rough” view about the nature of explanation which suffices for their purposes is simply to ignore these kinds of questions. Fourth, an open contempt towards philosophical arguments raises the question of how seriously philosophers should take the model of reduction that is being suggested. It is unclear whether philosophical arguments can move those who do not take these arguments seriously. Philosophers can point out the conceptual confusions into which scientists have fallen or are in danger of falling. A refusal to consider the relevance of philosophical arguments shows intellectual dishonesty. On the other hand, this refusal shows how wide the gap today is between philosophy and scientific philosophy.

Let us now turn to Davidson’s arguments for why mental–physical identities cannot be found, and to his reasons for asserting that our mental vocabulary is not an optional part of our conceptual resources. The fact that recent models of reduction do not consider these arguments shows a serious defect in their capacity to discuss the alleged irreducibility of the mental.
3.3.2 The subjective and intersubjective nature of thoughts

We have briefly discussed how the attribution of mental states is governed by normative principles and how the purpose of these attributions is meant to reveal the subject as a rational animal. According to Davidson, it is a condition of having thoughts that the basic standards of rationality have application. In giving reason-based explanations of actions, we give the agent's reasons for doing what he did. Following Davidson, we call these kinds of explanations rationalizations and say that the reason rationalizes the action. In rationalizations normative constraints apply because we have to consider what inconsistencies do least harm to intelligibility. The normativity of rationalizations is partly a result of the fact that the interpreter has to use his own standards of rationality when attributing mental states to others and these standards dictate, in a very authoritative way, the shape of the pattern in which propositional attitudes can be said to exist. As long as we want to use reason-based explanations, we have to find rationality in the target of attribution. Or to put it the other way, we can use reason-based explanations only to the extent that we find rationality in the target of attribution. There is no rationality without reasons and there are no reasons without rationality.

When discussing the anomalism of the mental, we already saw how the normativity and holism of the mental leads Davidson to argue that mental concepts belong to a different conceptual domain than physical concepts. The normative methodology of interpretation arises from the norms of the mental. Although both the physical and the mental are governed by norms, the nature of the norms and the way they are used in the case of the mental are different from those of the physical case. According to Davidson the disparate character of these norms explains the “ultimate... difference between understanding minds and understanding the world as physical.” He claims that in contrast to the study of the physical, norms enter in “... a special and additional way in the study of mental phenomena.”

It is only in Davidson’s later writings that the role of normative methodology and the reasons for it become clear. These reasons were already anticipated in his earlier writings, but important additional arguments were provided later. Davidson’s essential claim is: “The

---

1193 See Davidson, 1963.
1195 Davidson, 1990b, 98.
mental vocabulary... is irreducible because it is normative.”¹¹⁹⁶ It is clearly the normativity of our mental vocabulary which sets it fundamentally apart from the vocabularies of the natural sciences. This aspect is not appreciated by those who think that the mental–physical distinction does not differ from the distinctions between different physical vocabularies. As I have shown, this is apparent in the case of Bickle and Churchland, who dismiss Davidson’s arguments without further consideration. To some extent the same is true of Kim who observes:

[…] intentional psychology may differ from biology in a way that can make a difference: normativity is central and essential to…psychology but has no place in biology. What difference does this make to the scientific character of…psychology? This is a question worth pondering.¹¹⁹⁷

It is strange that Kim does not ponder this question further. The fact that Kim does not pay close attention to the arguments about the normativity of psychology is very curious because he has also noted:

The right way to save… psychology… is to stop of it as playing the same that ‘cognitive science’ is supposed to play – that is, stop thinking of it as a ‘theory’ whose primary reason d’etre is to generate law-based causal explanations and predictions. We will do better to focus on its normative role in the evaluations of actions and the formation of intentions and decisions.¹¹⁹⁸

The lesson about the normative role of psychology, the lesson of Davidson, von Wright and Stoutland that Kim suggests, is a valuable lesson to Bickle, to the Churchlands and to other reductivists. Davidson obviously thinks that the normative / non-normative distinction is the reason which distinguishes the two vocabularies and it is therefore no wonder that its importance is strongly emphasized. This being the case, it is surprising that critics neglect this argument.

Essential questions for the argument for irreducibility include ones about how normativity infiltrates the descriptions of mental phenomena and why this happens. The normativity of mental state attributions means that we have to use our own norms and understanding when addressing the question of what rationality consists in. The reason why we have to do this is that we cannot attribute mental states at all without assuming that the subject is rational. There are no specific norms of rationality, and in belief attributions we therefore employ our own norms; the only way to see whether these norms are shared by

¹¹⁹⁷ Kim, 2003, 133.
¹¹⁹⁸ Kim, 1989, 263, fn. 46.
someone else is to assume that in large part they are. Every case of norm-governed
interpretation is unique, and we may have to adjust our interpretation according to the
contextual factors present on each occasion. This means that interpretation is always an open-
ended project.

Perhaps the normative dimension of the mental cannot be clearly separated from
questions of how our attributions work and why they must work that way. A partial answer to
the “how” question is that we attribute mental states to others according to our own norms and
understandings of rationality. A partial answer to the “why” question is that we do this
because it is the only way to make the interprettee intelligible. This is the only way to
understand the subject. There is an irreducibly subjective element in interpretation because we
have to rely on our own norms and understandings when considering the question of how
intelligibility can be secured. But I believe that a more complete answer to the “why” question
clarifies in an important way the irreducibility of the mental. In an important passage
Davidson notes:

When what is studied is the mental, then the norms of the thing observed also enter. When thought
takes thought as subject matter, the observer can only identify what he is studying by finding it rational
– that is, in accord with his own standards of rationality. The astronomer and physicist are under no
compulsion to find black holes or quarks to be rational entities.\footnote{Davidson, 1990b, 98, my emphasis.}

This quote indicates that the fact that we have to use our norms when describing the mental
does not in itself take the mental to a conceptual level which is distinct from the physical. It is
rather the fact that the norms of the thing observed also enter the picture and we are forced to
consider them. When studying black holes we are not compelled to do anything; we just
observe. According to Davidson, here lies the ultimate difference between “psychology and
the rest”; what is special about psychology is that the objects of study are subjective, norm-
governed states and these states, in addition of being under the norms of the interpreter, are
also under the norms of the subject. The fact that there is a mind at both ends forces us to
consider how to best match the contents of two minds. Such considerations have no place in
the natural sciences, because when we treat the world as mindless nothing corresponds to the
normative and subjective dimension of the mental.

It is interesting that in his later writings Davidson is eager to point out that the
difference between psychology and the other sciences or between our intentional and non-
intentional vocabulary comes back just to this: in psychology the object of study is the mind.
It could be argued that this is utterly trivial; in fact what could be more obvious than the fact that the nature of conceptual or empirical studies is shaped by the subject matter? Well, if we accept the claim that the subject matter is the reason why a certain methodology, a certain vocabulary, is required, then we are a step closer to the view that there may be an irreducible difference in our *ways of studying* the mind and the world (the latter as mindless). Those reductionists who insist that “minds are just part of the nature”, and that minds could be reduced to molecules, do not consider what is distinctive about the mind as compared to other physical phenomena. The distinctive feature of interpretation and reason-based explanation is that I employ *my* thoughts and values to construct a picture of another person’s thoughts and values. The difference between these kinds of explanations and explanations in the natural sciences is thus the fact that the object of my description is itself a rational subject. Many seem not to have grasped the fundamental importance of this fact in Davidson’s argumentation. Replying to Rorty Davidson notes: “[…] interpreting others is a matter of using (not looking at) my own values and thoughts, my norms and rationality, to understand someone else’s. I do not expect to find propositional attitudes, or the kind of norms and rationality they entail, in a beetle.”

To those who appreciate the distinction but suggest that it can be overcome in the future when we know more of the physical nature of the mind, Davidson replies in a skeptical tone:

> I disagree with Ramberg on… one point. He believes it is possible that the difference between our mental vocabulary and our many vocabularies suited to describing and explaining the world in non-cognitive terms may shrink so much that we no longer “change the subject” as we shift from one to the other. I do not… What I am convinced of is that when we treat animals as rational, the norms of rationality are in play, and such norms have no role in our other ways of thinking.

Thirty years after “Mental Events”, Davidson still argues that shifting from a mental to non-mental vocabulary would be a change of subject. Davidson is also inclined towards the claim that our two essentially different modes of explanation and prediction are parts of a comprehensive theory, and this holds a promise of a possibility of conceptual unification. Davidson contests the idea of such unification and claims instead that the two modes of understanding are two different kinds of vocabularies, with neither being reducible to the other. The reason why Davidson contests the possibility of unification is precisely the rationality of minds. We do not expect that non-rational entities will suddenly turn out to be

---

1200 Davidson, 1999o, 600, my emphasis.
1201 Davidson, 1999a, 620.
1202 See Davidson, 1987a and Davidson, 1999n.
rational entities and that we could start employing our mental vocabulary with them. But what about the other side of the coin? Can we imagine a situation in which we, the humans living today, would turn out to be non-rational entities? Can we imagine a situation where our physical vocabulary could be applied to our mental lives? If the former possibility sounds absurd why would the latter possibility sound any less absurd? Our brain-cells are not rational, the molecules that make up those cells are not rational, even our brains are not rational. How could we describe our rationality in terms of these physical entities then?

We have a grasp of how normativity and rationality shape the way we explain the behavior of agents. When interpretation takes a normative tone, it involves subjective elements which do not shape our understanding of the physical. On the one hand, there is the subjectivity of the *interpretee*, but on the other hand there is also the subjective choice of the *interpreter*. Davidson notes:

> [...] an interpreter must separate meaning from opinion in part on normative grounds, by deciding what, from his point of view, maximizes intelligibility. In this endeavour, the interpreter has, of course, no other standards of rationality to fall back on than his own.  

Somewhat paradoxically, we cannot go outside our norms and check whether they are correct without assuming that they are. The only way that our own standards of rationality can be compared to the standards of others is to engage in a process of interpretation, which itself depends on the use of our own norms. The crucially important lesson is that whereas in the physical sciences we can agree on what the standards of judgments are, and in this sense agree on how an objective point of view can be reached, this cannot be done when interpreting others. Thus:

> We cannot in the same way [as in the natural sciences] agree on the structure of sentences or thoughts we use to chart the thoughts and meanings of others, for the attempt to reach such an agreement simply sends us back to the very process of interpretation on which all agreement depends.  

We cannot step outside interpretation in understanding others, in getting information about their mental states. But why not? Those who think that Davidson’s views imply instrumentalism or antirealism would ask this question. I think that Davidson is here raising a
very important point when he asks what the criteria for mental state attributions are. He is claiming that the interpretation of others forms the basis of our capability to understand anything. If you and have a different opinion, it is because we can understand each other; the grounds on which I can understand you cannot be derived from outside me. If I cannot make sense of you, then to whom may I turn for advice? I can always consult another person but this requires that I understand him and the process of interpretation continues. Davidson claims:

You and I cannot come to agree on the interpretation of our sentences as a preliminary to using them to interpret others, for the process of coming to such an agreement involves interpretation of the very sort we thought to prepare for. It makes no sense to ask for a common standard of interpretation, for mutual interpretation provides the only standard we have.1206

We have reached a situation where the irreducibility of the mental is no longer a matter of the “disparate commitments” of the mental and the physical. Instead we see how the irreducibility of mental concepts comes back to the fact that they are the base level upon which all agreement or disagreement depends. Intersubjectivity is required for thought. The possibility of communication and the knowledge of other minds – knowledge about the mental – form the basis of our concept of objectivity. There is no going outside this standard to check whether or not it is correct. There are application conditions for mental concepts which cannot be violated without changing the subject. Mental concepts, in turn, are a prerequisite for the concept of objectivity, which for Davidson is the mark of thought. To question the application conditions of mental concepts would, ultimately, require questioning our standard of objectivity. This cannot be done because this standard is the one against which all other standards are measured. The Wittgensteinian tone of Davidson’s position becomes clear from the following:

Understanding the mental states of others and understanding nature are cases where questions come to an end at different stages. How we measure physical quantities is decided intersubjectively. We cannot in the same way go behind our own ultimate norms of rationality in interpreting others.1207

There is a point where questions about the mental come to an end. In my opinion Crane’s claim that: “Davidson has to say that there is a point in our investigation of mental phenomena where nothing more can be said….1208 is therefore correct. But the reason that

1206 Davidson, 1997c, 83. See also, Davidson, 1998 and especially Davidson, 1991.
1207 Davidson, 1994a, 233.
1208 Crane, 2008, 80.
we have to stop asking further questions about the nature of the mental is more fundamental than what Crane suggests. It is not merely that there is a point beyond which nothing more can be said about mental–physical relations. It is rather that when an ultimate disagreement about mental facts arises, we cannot seek an answer from any other source than from ourselves; in the end everything must come back to interpretation.

As long as we think that mutual interpretation is the only way to find out what someone believes, we seem to be trapped in a situation where the only objectivity that we can get for our norms and standards of rationality come from us, and this in turn requires the knowledge of other minds. This problem is stressed on many occasions by Davidson: “The trouble with the study of thoughts is that the standards of rationality are not agreed upon. We cannot compare our standards with those of others without employing the very standards in question. This is a problem that does not arise when the subject matter is not psychological.”

The “subjectivity” of interpretation becomes evident when Davidson notes:

If we ask what the criteria are for saying that some object is three feet long, the criteria themselves are objective in the sense that we can agree with other people as to what the criteria are. When it comes to keeping track of what is in somebody else’s mind, there is no way to agree on criteria because our contacts with other minds were the basis of the criteria. So, when we ask ‘what does somebody think, believe, or want?’ all we can do is to relate their states of mind to our own states of mind. There is only a subjective yardstick…. This is what I take to be the deep difference between the social sciences and the physical sciences. There is a sense in which the yardstick we use is unshared when we talk about other people, whereas when we are talking about the outside world, it is shared.

This captures something of the logic of mental and physical concepts. These conceptual reasons are the most effective arguments against mental–physical reducibility. Needless to say, these reasons are usually dismissed by scientifically minded philosophers. Davidson argues that physical objects and events lie, so to speak, halfway between people and this makes them objective in a different sense than the concepts that apply to mental states and events. He notes: “[…] external objects are at an equal distance between us and we try to triangulate them. But if we ask ourselves about the type of communication between two minds, the feature is different.” In the mental case, “the measuring stick is not shared”. This comes close to von Wright’s observation that the criteria for mental state attributions differ from the criteria for physical states.

---

1209 Davidson, 2001f, 131.  
1210 Borradori, 1994, 51-52.  
1211 Borradori, 1994, 52.
An interesting question is how much the “subjectivity of the yardstick” depends on the fact that interpreter and interpretee have unique collections of mental states. Every object that is three meters long is similar to other objects that are three meters long in this sense: they share a physical property of being three meters long. How about a group of people who all believe that snakes are animals? It is not clear that all these people are similar to each other in any interesting respect. It is clear that that they need not share any physical property which would correlate to their respective beliefs about snakes. When we talk of people “having the same belief” or “sharing a belief”, we mean that they are in states that we can understand. The only way we can understand others is to evaluate their states of mind from the perspective of our own states.

The previous lengthy quote is not just a comment on the difference between the social and natural sciences. It is rather a more general observation that our understanding of other minds and our understanding of the physical world are different. Someone disagreeing with this claim does not grasp the basic idea that Davidson is suggesting, namely, that arriving at an understanding of a speaker’s language and thought involves a different kind of procedure than arriving at a theory in physics or chemistry. Only the former is called interpretation, and this task is governed by different principles than scientific investigation or explanation. This simply is Davidson’s reason for believing why the social sciences cannot be reduced to the natural ones, or why the notions essential to interpretation cannot be reduced to non-mental notions. We have discussed the explanatory functions that Davidson takes mental concepts to perform. The variety of these functions can be considered simply by considering how we use mental concepts in our life. Two features of mental concepts, which have no echo in physical theory, should be emphasized. First, in our mental reality, there is the idea that something ought to be in a certain way. Given what a subject thinks, certain implications should follow. In the physical realm there is no sense in which objects and situations ought to be one way rather than another. If we occasionally use the expression that things should be this or that way when describing physical reality, we are merely expressing our expectations about the nature of physical phenomena.

But the identity of physical objects and events is obviously not governed by normative principles. In mental reality, the distinction between is and ought is a constitutive distinction. The second aspect that is clearly essential to our mental vocabulary but which is absent in our physical vocabulary is the idea of reasonableness. It is hard even to try to describe the meaning of this term without using the mental vocabulary to which reasonableness is so closely tied. How do you explain in a physical vocabulary that “If S believes that P, it is
reasonable for him to believe that Q”? Reasonableness seems to be a crucial feature of rational explanations, because when giving these explanations we have to weigh evidence according to our own view concerning what seems reasonable. Both these features of mental concepts come back to the fact that by using these concepts we are trying to give a certain kind of explanation.

In what sense does the difference between interpretation and scientific investigation or explanation imply the irreducibility of the mental to the physical? It is interesting that the difference between these tasks is what, in Davidson’s view, rules out strict psycho-physical laws and in this respect the argument seems just to repeat the claim that the conditions of coherence, rationality and consistency have no echo in physical theory, and that bridge-laws can exist only between conceptual domains which are, in an abstract sense, similar enough. But I think that the more general point is that since any understanding of another mind differs in kind from the understanding of the mindless world, there is no way in which the former study could be conducted with a methodology that is appropriate in the latter case. It should be obvious why cases of interpretation differ from situations where we treat the world as mindless. Since normative considerations are absent in the case of, say, biology and geology, these sciences can be seen to belong to the same conceptual domain as physics, whether or not they in fact reduce to physics. Natural sciences treat the world as mindless, and this is what puts them on the same conceptual level with each other.

An interesting question is whether mental phenomena should be thought of as being essentially different than the phenomena described by the physical sciences. Davidson would deny this, because in his view mental phenomena are physical phenomena and a mental description of a phenomenon is just one description among others. Kim asks: “If beliefs are essentially normative and are posited because of our normative requirement, are there beliefs in the same sense in which there are physical objects and events, like trees and explosions?”\textsuperscript{1212} This question is obscure, however. What would it mean to say that there are beliefs “in the same sense” as there are trees? What would an affirmative answer require? I believe that we indeed have to conclude that mental phenomena are different than physical phenomena; there are no beliefs in the same sense in which there are physical objects. What von Wright has noted about the mode of being of mental phenomena should be taken seriously: they have no spatial location and include an irreducible subjective aspect.

\textsuperscript{1212} Kim, 2003, 134.
Because Davidson rejects mental antirealism, he would say that from an ontological point of view there is nothing second-grade about mental phenomena and that attributions of mental phenomena are as objective as the attribution of physical phenomena. But it seems to me that from what has been said above, we should conclude that there is something different, something peculiar, about mental states and therefore I believe that Davidson’s claim that “judgments concerning these [mental and physical] phenomena are true or false in the same way”\(^1\) must be modified. True, mental and physical statements are true “in the same way” in the sense that both are responsible to their respective regulative principles and must respect them. This, as we have seen, is one reason for the anomalism of the mental. But, as Davidson notes, “[…] there is difference in the nature of the case between the kind of knowledge we have of other minds, and the kind of knowledge we have of external objects.”\(^2\) I think this statement raises the question that Kim asks, and although the question is perhaps not as ontological as Kim thinks, it certainly seems that there is an essential difference concerning the attributions of mental and physical states. Whereas Davidson’s physicalism may lead us to think that the physical description of an event is privileged and, consequently, may lead to accusations of eliminativism, instrumentalism or antirealism with respect to Davidson’s position, I think that we should draw the very opposite conclusion about the status of Davidson’s position. Given that our understanding of other minds is the basis for our knowledge about other things, it follows that “if our judgments of the propositional attitudes of others are not objective, no judgments are, and the concept of objectivity has no application.”\(^3\) So, *pace* Bickle, the role of psychology is not to function as a servant for neuroscience. If we have to put psychology and other sciences or our mental and physical vocabularies in order in terms of their epistemological importance, it is psychology which leads the others.

The irreducibility that we are discussing here is best understood as methodological irreducibility. A conflict occurs if there is a methodological difference between the study of the mind and the mindless world, and if the reducibility of a theory to another is thought to imply that the use of the vocabulary of the reduced theory can be stopped. That mental concepts are norm-determined is a description of a feature of their application. There is a conviction, for which reasons have been offered, that we have to use two different kinds of vocabularies when describing and explaining reality, and Davidson is describing what the

\(^1\) Davidson, 1973, 254, my emphasis.  
\(^2\) Borradori, 1994, 54.  
\(^3\) Davidson, 1997c, 84.
essential characteristics of each vocabulary are and why these characteristics make the mental vocabulary irreducible. As I have emphasized, the irreducibility due to normativity has been left without discussion by many critics. “Irreducibility” in this context means “only” that with our mental vocabulary it is possible to give descriptions and explanations which cannot be given in our physical vocabulary. As long as these descriptions are useful, there is a reason to preserve our mental vocabulary and clarify its features and application conditions.

It could be argued that the irreducibility of psychology and mental concepts arises from a very practical source, namely, from our special interest in interpreting human agents as rational agents. We seem to have an interest, perhaps largely biological in nature, for the reasons for our actions and other psychological phenomena. Davidson claims:

The constitutive force in the realm of behavior derives from the need to view others, nearly enough, as like ourselves. As long as it is behavior and not something else we want to explain and describe, we must warp the evidence to fit this frame.... Standing ready, as we must, to adjust psychological terms to one set of standards and physical terms to another, we know that we cannot insist on a sharp and lawlike connection between them. The limit thus placed on the social sciences is not set by nature, but by us when we decide to view men as rational agents with goals and purposes, and as subject to moral evolution.1216

Davidson argues that the irreducibility of the mental is not set by nature but by us. In one of his first articles Davidson admits that we can choose to treat human beings under physical concepts if we wish.1217 We have seen that Bickle, for example, claims that Davidson’s argument is without force against new wave reductionism. But this claim neglects Davidson’s view about the “deep difference” between psychology and the physical sciences. Bickle’s interpretation is based on his reading that: “Anomalousness itself implies nothing about the explanatory power of intentional psychology compared with physical science. According to Davidson, neurobiological theories might prove to have complete explanatory and displacement potential....”1218 Bickle seems to recognize the Davidsonian claim that we can choose to treat humans under physical concepts if we wish. Bickle’s interpretation, however, suggests that intentional psychology could be compared to physical science, and in doing this we would come to realize that the latter is superior and has complete explanatory and displacement potential with respect to the former. This cannot be Davidson’s view, since he insists that we need two different modes of description, explanation, and prediction to understand reality.

1216 Davidson, 1974b, 239, my emphasis.
1217 Davidson, 1964.
It is interesting that although Bickle sees Davidson as a contemporary dualist, he argues that Davidson can be interpreted also as being a reductive physicalist or eliminativist. Bickle quotes Davidson’s following statement: “The nomological irreducibility of the psychological... does not mean that there are any events that are in themselves underdetermined or unpredictable... events described in appropriate physical terms, are as amenable to prediction and explanation as any.” He claims that this justifies the eliminativist interpretation. Bickle notes also that, according to Davidson’s monistic view, every mental event has also a physical description. Surprisingly, Bickle has failed to note Davidson’s conviction that “even if someone knew the entire physical history of the world, and every mental event were identical with a physical, it would not follow that he could predict or explain a single mental event....” If we recognize that a description formulated in our physical vocabulary is not a competitor to a description formulated in our mental vocabulary, then questions about explanatory and displacement potential seem to lose their importance and in fact, their very intelligibility.

3.3.3 The dispositional character of mental concepts

When discussing the reasons for the irreducibility of the mental, we must keep firmly in mind that the discussion of the previous and present sections is based on the view that reduction is a relation between linguistic categories, and not between ontological categories. If we think that the conceptual framework of folk-psychology or common sense is both useful and indispensable, then there is a rationale for studying the logic and principles of this framework and the reasons for its irreducibility and indispensability. The main motivation for defending our mental vocabulary comes from the conviction that we can do things with this mental vocabulary that we cannot do with our physical vocabulary. Because of their propositional contents, psychological concepts can be used in reason-based explanations. Given that there are these tasks which mental concepts are meant to perform, it is important to realize what it is about the concepts that enable us to do this. What kind of concepts are mental concepts?

According to Davidson, an essential feature of mental concepts is their holistic and normative nature. These features demand the use of interpretative principles which cannot be expressed in a vocabulary which is devoid of propositional content, and therefore devoid of the logical connections which relate mental phenomena to each other, and which matter for

---

1219 Ibid.
1220 Davidson, 1974b, 224.
psychological explanations. Another aspect of mental concepts that has a role in the discussion of the irreducibility of the mental is their dispositional character. Davidson refers to this aspect by claiming that mental concepts are *causal concepts*. This feature was not stressed in “Mental Events”, but is something which Davidson strongly emphasized in his later writings. Interestingly, Davidson notes that he has “almost from his first essay” emphasized the essential and ineliminable way in which causality is built into mental concepts and explanations. This is a very interesting remark for the following reason: Davidson is often seen as *the* key figure in restoring philosophers’ faith in the idea that reasons must be causes. The previous comment could be read as a suggestion along these lines, that is, as a suggestion that Davidson has since the beginning emphasized that mental concepts and explanations are causal. This “reductive” reading of Davidson’s position, which equates reasons with causes, has been the standard reading. But it seems clear to me that what Davidson is saying here means that he has from the beginning emphasized the ineliminable causality of mental concepts in the sense that this feature makes these concepts irreducible and irreducibly different from physical concepts.

In chapter two it was noted that it is part of Davidson’s definition of a strict law that it does not deploy dispositional terms or use causal concepts. Davidson claims that certain concepts cannot be clearly understood without appealing to causality. These concepts are *irreducibly causal*. Some examples are the concept of favism, of an infectious disease, or of sunburn. The first is the concept of an allergic reaction which is caused by the broad bean, the second is the concept of a disease which is either capable of causing infection or is caused by a micro-organism, the last is a state necessarily caused by the action of the sun. Davidson claims that the concept of acting with an intention and the concepts of belief and desire belong to these kinds of irreducibly causal concepts. The reason is that an intentional action is something which is caused, in the right ways, by beliefs and evaluative attitudes. Mental states in general are partly identified on the basis of their causes and effects.

Davidson claims, perhaps surprisingly, that getting rid of causal or dispositional concepts is a mark of *progress* in a science. The most developed level would be the level of “developed physics”, where all references to causal concepts would have been removed. Davidson claims that the promise that all reference to causal concepts can be dropped is “intrinsic to physics”. What is meant by the claim that science tries to dispense with causal concepts is that in a mature science we would not be satisfied if told, for example, that a glass

---

1221 Davidson, 1999y, 106.
1222 Davidson, 1997c).
broke because it was *fragile*. It is just when the relevant physical mechanisms are unknown that diseases are defined by their causes or symptoms and the breaking of a glass is explained by its frangibility. When the mechanism *is* known, the explanation of why a glass broke will not refer to the causal concept of frangibility. The concept “fragile” is no longer needed to explain the breaking of the glass. An appeal to causal dispositions leaves something unexplained, because it fails to tell what it is about objects that make them behave as they do. When we attribute a causal power to an object we say that there is *something* about it that will cause certain effects given the right conditions. This implies that a “deeper” answer is available, an answer which explains the causal properties of the object. In the physical sciences which approach, or at least try to approach, the ideal of explanatory unity, there is a promise that explanations can be replaced and sharpened without loss. If we can explain the mechanism behind favism, as I am sure we already have, and dismiss with the original concept, we have not lost anything relevant for explanatory purposes, *at least not from the perspective of physical science and from the perspective of causal explanations*. Davidson’s claim is that in the physical sciences there is no reason, save ignorance, why we could not substitute non-causal concepts for causal concepts. We should recall what has been noted already: sciences, or phenomena, which are not constrained by normative considerations, belong to the same conceptual domain.

Mental concepts are not like physical concepts because their causal aspect cannot be dismissed. It is misleading to say that ordinary common-sense physical concepts, like that of solubility, and our mental concepts, like that of belief, are irreducibly causal *in the same way*. It is true that ordinary physical concepts are identified in part also by the sorts of happenings they are prone to causing, given the right conditions. But in these cases there is a promise that a more detailed explanation for why certain things happened can be given. In the case of mental concepts their causal character cannot be dispensed with. Why not? Because: “[Mental concepts]... appeal to causality because they are designed, like the concept of causality itself, to single out from the totality of circumstances which conspire to cause a given event just those factors that satisfy some particular explanatory interest.” Mental concepts have this feature because rational explanations are interest-sensitive in a very strong sense. As Davidson claims: “When we want to explain an action... we want to know the agent’s reasons, so we can see for ourselves what it was about the action that appealed to the

---

1223 We can recall Bickle’s comment that there is no such thing as a “psychology of memory consolidation” because we already know, at a lower level, what memory consolidation *is*. When we know what “frangibility” is, not just functionally, but, say, on the molecular level, all explanations referring to “frangibility” become otiose.

agent.” We should consider what is being done with mental concepts and how this explains some of their features. Remove these features and the concepts cannot perform the tasks they were originally meant to perform any longer. We could claim that removing certain essential features of mental concepts would make them unsuitable for the language-game which was originally their home.

According to Davidson, both mental concepts as well as the concept of causality itself are “designed” to single out a factor or factors that answer to our explanatory interests. When we want to explain an action, we select, interpret, the agent’s reason(s) so that we can see what it was about the action that appealed to the agent – and by doing this, we rationalize the action. For this kind of explanation of action, we have to select the right conceptual connection that the agent saw, or thought he saw. Likewise, in the case of physical events, we often choose the cause in a way that depends on our explanatory interests. This means that, depending on the vocabulary we use to identify an event, different sorts of explanation are appropriate. The language of causality itself has no place in strict laws, because when we use the language of causality we select causal factors whereas strict laws don’t select; that is precisely what, in Davidson’s opinion, makes laws strict. Perhaps the view that mental concepts are designed to single out specific causal factors already explains why they cannot be reduced to strict laws. But there is an additional reason that prevents the reduction. It is the way that we choose what the specific causal factor is. Namely:

[...] beliefs, which are also causal dispositions, are specified in terms of their relations to one another and to the events and objects in the world, and, in judging the relevance of these relations to the identification of particular beliefs, norms are necessarily employed. In order to keep intact the normative features that help define beliefs and other thoughts, a degree of looseness in their connections with events as described in non-cognitive terms is required. Not surprisingly, it is again the normative dimension of the mental which makes the causality of mental concepts a special case. The holism and externalism of mental states are used to defend the claim that mental concepts are, so to speak, “doubly-causal”, since they are identified both in relation to each other and with respect to the outside world. These relations cannot be observed to hold without the infiltration of some normative methodology. As Davidson notes, the normative features must be kept “intact”, which is to repeat the claim that in belief attribution we cannot step outside our conceptual framework of the mental, which does not share its constitutive features with the physical. In the case of the mental, causality is

---

1225 Ibid.
1226 Davidson, 1990b, 98, my emphasis.
connected with the normative demands of rationality; causality enters the picture in a very different way than in cases where we describe physical objects and events in terms of causal concepts. Perhaps it could be said that in the case of physical events and objects one can, in a sense, see that there is a causal relation between physical phenomena. It is of course an old idea that one cannot really see the relation of causality itself, the relation is inferred. But consider the paradigmatic example, a billiard ball B hitting ball B₁ thus causing B₁ to move. In the colloquial language we certainly say that we saw that B caused B₁’s movement. We cannot in the same way say that we saw a belief causing an action or that we saw how the belief caused an action. I suspect that a large part of the problem of mental causation is a result of the fact that our concept of causality is applicable, first and foremost, to physical objects and events, to those objects and events that can be publicly observed. Therefore the problem of mental causation is, above all else, a result of a conceptual confusion. The concept of causation is from the physical realm and we attempt to apply it to mental reality; this results in confusion.

Strict laws do not employ causal concepts, and since mental concepts are irreducibly causal they cannot occur in strict laws. An interesting question is whether mental concepts are unsuitable for strict laws because they are irreducibly causal concepts or whether they are causal concepts because they are irreducible. Should the causal nature of mental concepts be used to defend their irreducibility to strict laws, or should we conclude that these concepts operate in a different way than physical concepts precisely because they cannot be incorporated into the realm of physical concepts? Is there a reason to argue that the situation could be seen in both ways? Davidson notes: “[...] belief, desire, intention and action... are concepts that cannot, without losing the explanatory power they have which binds us to them, be reduced to the concepts of an all-encompassing physics.”¹²²⁷ He claims also that “the fact that [mental] concepts are causal concepts makes those concepts ineligible for inclusion in strict laws.”¹²²⁸ Here the claim is clearly that mental concepts, because they are causal, cannot be used in strict laws. Their special explanatory power, due the fact that mental concepts are causal concepts, prevents their reduction to physics. Psychology concentrates on the causal role of reasons. But this view should be compared to the following: “I have been suggesting that appeal to causal concepts is appropriate to the explanation of action in part precisely because strict laws are not available”.¹²²⁹ Here the suggestion seems to be that since

¹²²⁷ Davidson, 1995c, 216.
¹²²⁸ Davidson, 1993e, 312.
there are no strict laws of action, or more generally, strict laws involving the mental, the idea of causal powers, of certain things making other things happen, has survived in the case of mental, whereas this idea becomes extinct in physics. If appropriate laws were available, there would be no need to describe causes in terms of the effects they tended to produce.

It seems that with respect to this question Davidson has gone through a change of mind, or he has at least later clarified his earlier views to a notable extent. In 1973 Davidson writes: “Unavoidable mention of causality is a cloak for ignorance; we must appeal to the notion of cause when we lack detailed and accurate laws. In the analysis of action, mention of causality takes up some of the slack between analysis and science.” From this statement we get the impression that an appeal to the concept of cause is necessary because we are ignorant of the strict laws that back up the causal relation. This, of course, relates to Davidson’s view that each singular causal relation must be covered by some strict law. In the analysis of action, the law would be expressed in terms of our physical vocabulary, and ultimately, in terms of our finished physics. But we do not have knowledge of the relevant law; we are ignorant about it. However, in 1990 Davidson writes: “The ‘unscientific’ concept of cause takes up the slack [between beliefs and events described in non-cognitive terms]. This slack is not the slack of ignorance: it is the slack that must exist between two schemes of description and explanation, one, the mental, being essentially normative, the other not.”

I think we can conclude that also the argument against reducibility, which is based on the dispositional character of mental concepts, comes back to the view that mental concepts serve different purposes than physical concepts – and this is what creates the real slack between the two vocabularies. What in my opinion is essential is Davidson’s later view that this slack does not occur because of ignorance, it occurs because there cannot be strict psycho-physical laws and thus some slack between the mental and physical must exist.

Here we encounter another reason why strict psycho-physical laws are impossible: if mental concepts were reduced to physical concepts, they would lose their essential causal character (which connects with the normative demands of rationality) – and this reduction would, once again, amount to changing the subject from the mental to the physical. Davidson’s change of mind about the dispositional nature of mental concepts is important for the following reason: if we are under the assumption that the mention of causality is due to ignorance, we may be tempted to conclude that this ignorance may be cleared once our scientific knowledge increases. This, in essential respects, is how eliminativists or reductivists

1230 Davidson, 1973b, 80.
1231 Davidson, 1990b, 98, my emphasis.
think. They claim that the explanatory power of mental explanations is not special and will eventually be explained in terms of physical mechanisms; this will make the former explanations redundant. If the explanatory power of mental explanations will not be explained in terms of physical mechanisms, the explanations will be considered pseudo-explanations and will ultimately be eliminated. If one is not willing to accept the radical Davidsonian conclusion, one is perhaps tempted to think that the irreducibility of the mental is just a practical matter which troubles us now, but which can be solved in the future. This would mean that the irreducibility of the mental is not principled. Davidson’s reformulated claim that the slack between the mental and the physical must prevail dismisses the possibility that the slack could be removed as science progresses.\(^{1232}\)

3.3.3.1 AM’s relation to the functional model of reduction

In section 3.3.1.1, the question of whether new wave reductionism can avoid Davidson’s challenge was considered. The conclusion was that Davidson’s views about the irreducibility of the mental have not been fully addressed by Bickle. Because the model of new wave reduction is still programmatic, the final verdict about its relation to the Davidsonian view cannot be fully stated. In this section I claim that also another new form of reduction, Kim’s model of functional reduction, can be challenged by Davidson’s arguments. Kim recognizes this.\(^{1233}\) He leaves Davidson’s arguments without discussion. My purpose is to consider how the model of functional reduction could be evaluated from a “Davidsonian perspective”. Kim offers his model of reduction as the best, and perhaps only, alternative for saving physicalism, and therefore its success is relevant for those who doubt the prospects of physicalism. Evaluating how the “best” model of reduction relates to Davidson’s arguments against the reducibility of the mental is also helpful in deciding whether Davidsonian views are still relevant. A thorough analysis of Kim’s position cannot be done here, so my exposition will be somewhat cursory and tailored to the above purpose.

As I showed in the previous section, Davidson argues that mental concepts are irreducibly causal. This claim has certain similarities to Kim’s view that mental concepts are functional concepts. The basic idea of functionalism is that mental states are constituted by their causal relations to one another and to sensory inputs and behavioral outputs. AM’s

\(^{1232}\) It is not clear whether Davidson changed his mind. If not, the later way to put the issue is more illuminating. Davidson (1993e, 312) himself notes that the previous formulation was “misleading”.

\(^{1233}\) See Kim, 1998, 93 fn. 7.
relation to functionalism is an interesting question.\textsuperscript{1234} Notwithstanding Davidson’s relation to the “program of functionalism”, he nevertheless thinks of mental states in roughly the same way as many functionalists do. Although Davidson does not think of mental states as internal entities mediating between input and behavioral output, he thinks that mental states are specified in terms of their relations to each other and to events and objects in the world. From this an interesting question follows. Kim has claimed that:

\begin{quote}
If the functionalist conception of the mental is correct… then mind-body reduction is in principle possible… This is contrary to one piece of current philosophical wisdom, the claim that functionalism… is a form… of mind–body antireductionism. What I am urging… is the exact opposite – that the functionalist conception of mental properties is required for mind–body reduction.\textsuperscript{1235}
\end{quote}

The interesting question is this. Davidson insists that mental concepts are causal concepts and Kim insists that mental concepts should be defined in terms of their causal roles if mind–body reduction is to be a possibility. Does Davidson’s understanding of mental concepts therefore actually lead to the possibility of their reducibility?

Kim’s occasional claims about the nature of mental properties actually come quite close to Davidson’s view. First of all, Kim suggests that we should stop thinking of mental properties as higher level properties and should instead formulate the discussion in terms of higher level concepts. This is surprising given that Kim insists that the problem of mental causation is, above else, the problem of how the “mental qua mental” can be causally efficacious. We have also seen that Kim wants to address the “robustly metaphysical” problem of mental properties. Sometimes Kim however argues that we should give up the talk of second-order properties in favor of second-order concepts. As far as I can see, this is essentially the Davidsonian view. As I have shown, in his view the mental is a conceptual category and Davidson wants to have the discussion in terms of mental predicates and not in terms of mental properties. In fact, the following statement of Kim could be from Davidson:

\begin{quote}
The use of second-order property designators probably is unavoidable, and we should recognize that these designators introduce a set of useful and practically indispensable concepts that group first-order properties in ways that are essential for descriptive and communicative purposes.\textsuperscript{1236}
\end{quote}

\textsuperscript{1234} For considerations about AM’s relation to functionalism, see Mcdowell, 1985. We have already seen that Davidson’s arguments for the irreducibility of the mental are essentially different than the standard argument of functionalism, which refers to multiple realizability. Therefore, as I noted in section 2.1.1, AM and functionalism are often referred as two different arguments against type-physicalism. As Kim (2003) correctly notes, Davidson’s antireductionist arguments differ from functionalism in an important respect, namely, whereas functionalism allows local reductions, the anomalism of the mental disallows even one-way laws like the ones that are supposed to connect mental states to their physical realizers. In fact, it seems to me that it is questionable whether functionalism should be considered as a form of non-reductive physicalism.

\textsuperscript{1235} Kim, 1998, 101.

\textsuperscript{1236} Kim, 1998, 105.
This view comes actually very close to Quine’s view. When Quine accepted AM he claimed that mental concepts are practically indispensable and insisted that we have to settle for a dualism of concepts. The only thing a “Davidsonian” would change from the previous quote is the term “probably” and would argue that, in the case of the mental, the use of second-order property designators is necessary. It seems to me that Kim’s view here is in essential respects similar to the view that I briefly considered and suggested as an answer to the question of whether mental properties are physical properties.\textsuperscript{1237} The view is that “the concepts introduced by second-order designators pick out first-order properties disjunctively”.\textsuperscript{1238} So far there is no fatal disagreement between Kim and Davidson.\textsuperscript{1239} Moreover, although Kim does mention Davidson in this context, the suggestion that multiply realizable properties, for example mental properties, are causally and nomologically heterogenous kinds is basically the view which was suggested already in “Mental Events”. Consider Kim’s claim:

> What lends unity to the talk of dormitivity and such is conceptual unity, not the unity of some underlying property. \textit{Qua} property, dormitivity is heterogenous and disjunctive, and it lacks the kind of causal homogeneity and projectibility that we demand from kinds and properties useful in formulating laws and explanations. But dormitivity may well serve important conceptual and epistemic needs, by grouping properties that share features of interest to us in a given context of inquiry.\textsuperscript{1240}

Given the view that was suggested in section 2.3.6 we should agree with this proposal; the quote seems to also agree with the spirit of Davidson’s argument. But Kim also thinks that mental states are \textit{locally reducible} to physical states. Should we agree with this? Accepting the claim that local reducibility of mental is possible surely seems to undermine the strong commitment to the irreducibility of the mental. It could be claimed that Davidson is dogmatic about the irreducibility of the mental when he claims that this irreducibility is principled. On the other hand, he has also discussed the possibility of local token-reductions. Perhaps the reducibility of the mental is a possibility if the requirements of reduction are loosened. We have seen that Davidson has very high standards of reducibility in mind, and this being the case it is possible that reduction with looser requirements could be accepted from Davidson’s

\textsuperscript{1237} See section 2.3.6.
\textsuperscript{1238} Kim, 1998, 105.
\textsuperscript{1239} Although I do recognize the possibility that if one interprets Davidson as a nominalist he would perhaps dismiss the suggestion that mental properties are physical properties. But as we have seen, Davidson does not seem to be uncomfortable speaking of \textit{physical} properties.
\textsuperscript{1240} Kim, 1998, 110.
Therefore, we should not dismiss the possibility of functional reduction straightforwardly but consider its plausibility by evaluating Kim’s view.

In the following I will briefly consider how Kim’s functional model of reduction is supposed to work. Let us suppose that we want to reduce a mental property M to a base domain of properties B. The first step is to functionalize M, by construing it as a property defined by its causal / nomic relations to other properties. We construe M as a second-order property, which is defined by its causal role. We give a causal specification H which describes the typical causes and effects of M. As a result, we understand M as a property with certain causal potentials. What is important is that the functionalization of M requires a functional definition of M, which takes the following form:

Having $M = \text{def.} \text{ having some property or other } P \text{ (in the reduction base domain) such that } P \text{ performs causal task } C$.

The property, for example, of being a gene is a second order property or second order concept. We define it as being a mechanism that encodes and transmits genetic information. This is the causal task C. On this model, being a gene (property M) is defined in terms of its causal task (C) and, by definition, P is the property which performs this causal task. Functionalization of M thus requires conceptual connections – definitions – which provide conceptual or semantic relations between the phenomena at higher and lower levels. The second step is to find the realizer(s) of M in the base domain B, that is, to find those properties or mechanisms in B that perform the task C. This means that in order to reduce M we have to identify it with its realizer P. The third step is to construct a theory explaining how P performs the causal task C. This model involving the three steps is precisely Kim’s functional model of reduction. It consists of: 1) functionalizing M; 2) finding the realizers of M in B; and 3) constructing a theory about how the realizer performs the causal / functional task which is assigned to it in the first step.

All of this looks good in theory and on paper. But I think that the explanatory answers that Kim obtains sound too good. For example, to the question “Why is it that whenever a

---

1241 Davidson notes (1987a, 111): “I was interested in the question whether reason-explanations are or ever could be just like the best explanations for which physics strives.” A reason why Kim does not consider Davidson’s arguments is precisely that, according to Kim, Davidson’s requirements for reduction are unrealistic.

1242 The details can be found in Kim, 1998, 1999, and 2005.

1243 As was noted earlier it is perhaps better to think of M as second-order property designator or second-order concept than as a second order property.

property P is realized in a system s, it instantiates a mental property M?” Kim answers: “[…]

by definition, having M is having a property with causal specification D, and in systems like
s, P is the property (or one of the properties) meeting specification D. For systems like s, then
having M consists in having P.”

Why does the system s instantiate M and not M* whenever it instantiates P? Because in systems like s, P is a realizer of M and not of M*.

These are satisfying answers according to Kim. But are they really? Obviously, it is possible
to state by definition that having M is having a physical property, which performs such and such functions – and that there is a certain low-level property which performs this function. But what does this definition tell us without actual functionalizations? Does it tell us anything more than von Wright’s observation that token physicalism can be taken to be trivially true? Without actual functionalizations, Kim’s model says very little about the actual connections between M and P properties.

Indeed, as Kim himself notes, “definitions are free” – if we please, we can define M as “having some property or other P such that P performs causal task C.” But what does this tell us about the prospects of the actual reducibility or irreducibility of the mental? A wise move on Kim’s part is the following: In “Having M = \text{def.} having some property satisfying causal role C”, we do not refer to a property M. If M here refers to anything, it refers to the concept or term M. The definition gives information about the concept M or about the meaning of M. This is a clever move because it allows Kim to say that such a definition in reductive explanations does not violate the following constraint on reductive explanations:

\begin{enumerate}
  \item[(R)] The explanatory premises of a reductive explanation of a phenomenon involving property F (e.g., an explanation of why F is instantiated on this occasion) must not refer to F.
\end{enumerate}

If we want to reductively explain M (“F” in the above constraint), we have to explain it without invoking the property M or other properties at the level of M. Otherwise the

---

1246 We should, of course, acknowledge that Kim’s model of reduction is a philosophical suggestion which is meant to give one possible picture of the relation between mind and matter. Kim argues that the relation must be something like this if we would like to give a reductive answer to the problem of mental causation. This is actually Kim’s whole point in his 1998 and 2005. But it seems to me that without actual reductions, there is no way to defend Kim’s model as an answer to the question of whether M properties really do reduce to P properties. Kim’s philosophical point is that given certain assumptions of physicalism, local reductions must be possible. This is an interesting conceptual argument, but its merit depends on the question of whether other physicalistic principles are accepted.
1247 Kim, 2005, 105.
explanation would not be reductive – reduction would not be achieved. So, why does x (a person or a system) have M at time t? We have the following kind of explanation (E):

\[
x \text{ has } P \text{ at } t. \\
P \text{ satisfies causal role } C \text{ (in systems like } x). \\
\text{Having } M = \text{def.} \text{ having some property satisfying causal role } C. \\
\text{Therefore, } x \text{ has } M \text{ at } t.
\]

According to Kim, the third line in the explanation does not refer to property M, but to the concept M. In order to reductively explain why x has M, we must formulate a definition like the one given on the third line in the explanation.

Let us suppose that we want to know why “Jones is in M” where M refers to a certain mental state. Would the kind of explanation given above be a satisfying answer? In some sense it could be claimed, yes. Let us suppose that we have defined M functionally and we have found its realizers. Then we could say that Jones is in M because he is in P. It is, however, unclear how well or whether this would satisfy our epistemic interests on the occasion – but we can grant that the explanation would at least give one answer to the question of “Why Jones is in M.” If this kind of explanation were available, perhaps that would be all we would want from a physicalistic perspective. The answer “M is there because P is there” is all that a physicalist studying the brain would want to know. This explanation would be a contribution to physicalism. However, even if we grant this, I believe there are reasons to be dissatisfied with the functional model of reduction.

First, how should we functionalize specific beliefs? How do you individuate beliefs? Whereas “gene” can be defined as “the mechanism that encodes and transmits genetic information”, what should we say of a belief that “Snow is white”? It is hard to think how we might start functionalizing it. If the functionalizability of specific beliefs is not required, then what grounds have we for the claim that we have “reduced minds”? Kim recognizes this problem of functionalizability, but comments:

[…] partial functional analyses of these [mental] properties… can get us going with the scientific projects of searching for the underlying physical / biological mechanisms. We don’t have to know all the things that belief does before we start work on uncovering its possible neural mechanisms…\textsuperscript{1248}

\textsuperscript{1248} 2005, 167.
But this view, as with so many other physicalistic assumptions, seems to be based on faith. Why suppose that partial functional analyses are available and why suppose that these would get us going with the scientific projects? What is a partial functional analysis of jealousy or love? Kim does not give answers. He does not suggest how we might actually functionalize mental properties, and the claim is basically that he does not see principled obstacles for functionalization and that he remains unconvinced by arguments to the contrary.\

Second, a closely related question is that of how to find the realizers of mental properties. Kim’s only suggestion in this regard is that this is the task of the scientists. Whereas Kim is carrying on a philosophical project, he gladly seems to hand the final task of reduction to scientists. But this proposal simply ignores the arguments of von Wright, Davidson, Malcolm and Wittgenstein. How can the realizers of mental phenomena be found if the very idea of token identity is confused? Third, Kim is almost completely silent about the question of whether Davidsonian arguments against the reducibility of the mental could be relevant when considering the plausibility of functional reduction. There are reasons to think that they could be relevant, and in fact there are a few hints which suggest that perhaps, despite his dismissal of Davidson’s argument based on the non-availability of strict laws, Kim thinks so too. We should remember that the question here is whether psychology, the concepts it uses, or the phenomena to which it refers, reduce to lower level sciences. Kim notes:

Davidson’s argument depends on mental anomalism, which in turn depends on the supposed special character of mental phenomena… namely their normativity and rationality. For this reason there is no reason to think that his argument can be generalized outside the mental domain.\

This is a very strange comment given that the issue is precisely one of the reducibility or irreducibility of mental domain! On the one hand, Kim says that if psychology is reducible by the same standards that apply to the best cases of theory reduction in the sciences, would it not be enough? But on the other hand, as we have seen, Kim himself raises the problem that psychology may differ from biology because of the normative role of the former and notes that this question should be considered in the discussion about the reducibility of the mental. Unfortunately he does not consider it. However, Kim’s comment that he has not come across “a totally convincing refutation of Davidson’s argument” suggests that an argument based

---

1249 Perhaps Kim would say that the actual functionalization is not a philosopher’s task. This could be a plausible reply but without a suggestion about how mental properties could be functionalized, is the model offered anything more than a proposal that we reduce mental states to physical states by showing what the physical counterparts of mental states are? This sounds trivial.

1250 Kim, 1998, 92.

1251 Kim, 1993a, xiii.
on the normativity of the mental could, also according to Kim, be relevant for the question the (ir)reducibility of the mental and the reducibility of the physical. From a personal correspondence with Kim I got the impression that in his view a functional reduction could suffice to explain the causal powers of mental states, but it would not completely reduce the mental because functional reduction misses the normative aspect of mental states. This answer, insofar as it admits that the normative aspect of mental is irreducible to the physical, would satisfy a Davidsonian, who would not insist on anything more.

A fourth problem which connects with the previous question is this. Although Davidson thinks that beliefs are specified in terms of their relations to each other, he insists that in judging the relevance of these relations norms are necessarily employed. In section 3.3.2 we saw how this should be understood; there is, in a sense, an irreducibly subjective aspect in interpretation that cannot be eliminated. Any attempt to eliminate it is itself an act of interpretation. It seems that there is therefore a very general question as to whether, even if a functionalization of mental properties were possible, the identities so established would be subordinate to our normal attributions of mental properties. If, as von Wright claims, the behavioral level has a semantic priority over the mental and the mental level has an epistemic priority over the neural, then it seems to follow that explanations like (E) are secondary when compared to our ordinary ways of establishing whether or why “Jones is in M at t”. I think that although (E) gives one kind of answer to the question of “Why Jones is in M”, the answer in terms of mental concepts would, because of the epistemic priority of the mental, be more basic than (E). It is not clear that Kim would have to absolutely oppose this, because he does not draw eliminativist consequences from the possibility of reductions. To grant the epistemic priority of the mental is to grant the irreducibility or autonomy of the mental.

The discussion here may give the impression that I associate Davidson’s views with functionalism and that a functional model of reduction looks plausible. With respect to both, the opposite is true. I think it is unclear what a philosopher claims when he says that mental states are “functional states”. A common understanding is that what makes something a mental state is the role which it has in the system of which it is a part. The identity of a mental state is to be determined by its causal relations to sensory stimulations, to other mental states, and to consequent behavior. But consider how we attribute beliefs to Georg. Let us suppose that we do it more or less in the Davidsonian way. Davidson, von Wright and Kim agree that intentional states are supervenient on behavioral and other observable physical facts. As a result of observing Georg’s behavior, we attribute to him the belief that “It is going to rain.”

The important question which we have already considered is: what kind of an entity are we
attributing to Jones? This was the question of section 2.5, and the conclusion was that von Wright and Davidson do not think of mental states as internal entities. This being the case, they would say that an attribution of mental states is not an attribution of brain states to the subject. The further questions which may arise from the attributions of mental states are secondary to the insight that these states are necessarily related to behavior. The necessity should be understood in the light of von Wright’s claim that behavior is the criterion of the mental and that the nature of the “purely mental” can be only be grasped through a counterfactual move. The relation between mental and neural states is accidental, whereas the relation between behavior and the mental is not.

One could understand functionalism also in a way which does not imply that a mental state is a brain state mediating between inputs and outputs; perhaps the state could be understood as a state of the whole person. Putnam, a former defender of functionalism asks: “Does it… make sense to suppose that what I am doing when I ascribe a belief… to someone is engaging in a bit of proto-scientific speculation about the internal causes of their behavior?”\(^{1252}\) According to Putnam, the functionalist conception of a psychological state accepted by Kim posits the existence of psychological theory which treats psychological states as theoretical entities which are to be identified with internal physical states of the subject. It is claimed that common-sense psychological concepts refer to these theoretical entities and are meant to give causal explanations of behavior in terms of functionally characterized internal states. Like Putnam, interpretationists would say that in attributing mental states we are not engaging in proto-scientific speculations about internal causes. As Child notes, “interpretationist [claims] that part of what it is for a creature to have a given attitude is for it to have a certain rationally specified disposition.”\(^{1253}\) As we saw in sections 3.3.2 and 3.3.3, this is Davidson’s view about attitudes. But this view, as Child also notes:

\[\ldots\] is much weaker than common-sense functionalism... for... [common sense functionalism]... requires, not just that a creature be organized in such a way that it has the relevant dispositional property, but that the psychologically specified causal roles be played by internal states or properties naturally identifiable in non-psychological terms.\(^{1254}\)

If we agree with von Wright and Davidson that the supervenience relation between the mental and the physical is a semantic relation, then there is little reason to think that that a similarity

\(^{1252}\) Putnam, 1999, 114.

\(^{1253}\) Child, 1994, 80, my emphasis.

\(^{1254}\) Ibid.
at the mental level implies a similarity at the physical level.\textsuperscript{1255} Whereas it is conceivable in the context of O-physicalism to think that if Donald and Georg are identical physically then they are identical mentally, what reasons do we have for thinking that “mental identity” would imply physical identity? Kim’s model of functional reduction requires the existence of species-specific or structure-specific bridge-laws. But why accept the existence of even such laws, i.e. why think that if I now believe that Q (I thus have the property M) and “having M = \text{def.} \text{having some property satisfying causal role C}” and given that P satisfies this role C, it would be the case that if I believed Q, \textit{tomorrow} it would be \textit{because} I am in P? Do structure-specific laws apply for future cases? Kim’s view about the nature of these laws is not entirely clear. He insists that, at the very least, there have to be local reductions even if they are only \textit{for single individuals at a particular moment} of their lives.\textsuperscript{1256} But would anyone, except a strong substance dualist, deny this?\textsuperscript{1257}

Let us suppose, to continue our example, that we have attributed the belief “It is going to rain” to Georg. Let us in addition suppose, in favor of Kim and \textit{pace} the interpretationists, that this belief is a certain physical property P of Georg. For reducing the mental property of Georg to a physical property, the former must be functionalized. As we have seen, Kim often uses the properties of “being a gene” and “being in pain” as examples of properties which have been functionalized. Let us not dispute the case of “being a gene”, which is a higher order physical property. But what can be said about pain? Kim suggests that “for an organism to be in pain is for it to be in some internal state that is typically caused by tissue damage and that typically causes groans, winces, and other characteristic pain behavior.”\textsuperscript{1258} But \textit{obviously} an organism could be in pain without suffering any tissue damage, as the phenomenon of phantom pain shows. An organism can be in pain even though it does not groan and wince; I could just assert very calmly that I am in pain. I can be truly in pain without suffering tissue damage or without wincing and groaning; we have agreed that there is an irreducible subjective element related to pain which cannot be captured in terms of behavioral criteria. If being in pain is “typically” caused by certain factors or is “apt” to cause certain reactions, what does this definition tell us about the question of whether pain is reducible to property P? It is interesting that, according to Kim, a functional definition of property M is actually a matter of establishing a \textit{conceptual or definitional} connection between M and the causal role

\textsuperscript{1255} See section 2.3.6
\textsuperscript{1256} Kim (2005, 25) notes: “[…] individual pains must… reduce to their respective neural / physical realizers.”
\textsuperscript{1257} The problem comes back to the question of what kind of entities mental states are. I may be perfectly entitled to ascribe the belief “Paris is the capital of France” to Donald without there being anything which is this belief in the brain of Donald.
\textsuperscript{1258} Kim, 1998, 19.
that it is meant to perform. He claims that conceptual decisions are often based on empirical knowledge of the relations in which M is embedded, and in practice the boundary between what is conceptual and what is not is going to be vague and shifting.\footnote{See Kim, 1999.}

We cannot specify what it is to believe that it is going to rain; how could we define a functional role which could be associated with this belief? We can roughly specify in mental terms a causal role characteristically played by a certain type of belief. But this specification is guided by our considerations of what is or would be rational for the interpreter to believe, and the specification is thus guided by the norms of the interpreter. Whereas Kim says that conceptual decisions are often based on empirical knowledge, Davidson thinks otherwise. According to him attitudes are dispositions; but they are not merely physical dispositions, they are also rational dispositions. This entails a conceptual connection between having an attitude and having a tendency to act in certain ways. But since the ascription of attitudes is constrained by normative considerations, the question of what attitudes a creature has cannot be wholly an empirical question. Nothing can count as a belief if it does not meet the standards of rationality. When Kim says that the nomic/ causal involvement of M is defining or constitutive of M, he is claiming that we can empirically find out what kind of relations M has to other things. But how could we? We could observe how the belief that it is going to rain manifests itself in the behavior of Georg and how it is related to his other beliefs, but in observing this we would have to evaluate the relevance of these relations from our point of view according to the norms of rationality. This being the case, the functional definition of M would actually be an artifact of interpretation and it would not, pace Kim, be based on empirical knowledge. This fact sends us back to the situation where the mental level, with its norms, is always privileged over the functional definitions that are a result of interpretation at the mental level. This is the point that, in their respective ways, both Davidson and von Wright make.

\section{3.4.4 Externalism}

One more reason for the irreducibility of the mental, which despite its importance does not receive much discussion in Davidson’s writings, is externalism. The alleged externalism of mental states is an interesting argument against the reducibility of the mental, because it perhaps tells us something about states of minds and not just about the ways we talk about
these states. The irreducibility which follows from externalism moves the discussion from the conceptual to a more ontological level. This is perhaps not the way that Davidson would have described the situation, since for him the mental constitutes a conceptual and not an ontological category. However, as I have argued, the question of the sense and extent to which mental states can be reduced to brain states may be interesting. *Can beliefs be found inside the brain? Where in physical reality are my beliefs?* Some philosophers think that these are not especially interesting questions. There are good reasons to assume that most cognitive scientists or neurophilosophers think otherwise. I think that an answer to this question has *some* relevance to *some* questions in the philosophy of mind and for sciences studying the mind.

Let us suppose that I have a belief and you have attributed the belief to me by familiar means. If externalism is true, it seems that we have no other way than the familiar way to make mental state attributions; we cannot find states of mind just by looking into the brain.\(^{1260}\) This in turn would mean that states of the mind are not intrinsic states of an agent, or at least that they cannot be identified by studying the internal constitution of the agent alone. Davidson discusses externalism and its relation to the irreducibility of the mental in two ways. On the one hand, the claim about externalism seems to be an epistemological claim, but on the other hand at times Davidson seems to be making more of an ontological claim. It is surprising that only few commentators have discussed this aspect of the irreducibility of the mental.\(^{1261}\) It is possible that most commentators equate the dispositional character of mental concepts with the externalism of mental states.\(^{1262}\) What may cause confusion is Davidson’s statement that one reason for the irreducibility of the mental is the “causal character of mental concepts”.\(^{1263}\) By this expression Davidson is however *not* referring to the argument described in section 3.3.3. After discussing the dispositional character of mental concepts, Davidson notes that:

\(^{1260}\) This is especially true with respect to beliefs; the situation is more complex with respects to beliefs. If O-physicalism is true, they should literally be inner states of an agent.

\(^{1261}\) The few exceptions are Shea, 2003, Rowlands, 1990 and Dardis, 2008.

\(^{1262}\) At times Davidson seems not to make a very clear distinction, as far as the irreducibility of the mental is concerned, between the dispositional character and externalism of mental concepts. He claims, for example, that: “[...] states of mind... are identified in part of their social and historical context in which they are acquired; in this respect they are like other states that are identified by their causes, such as suffering from snow blindness or favism...” (Davidson, 1988a, 51) In section 3.3.3 we saw how favism was used as an example of a causal concept, and the claim was that one reason for the irreducibility of the mental is the fact that mental concepts are also causal concepts.

\(^{1263}\) See Davidson, 1995a.
There is one more... consideration which militates against the nomological and definitional reduction of mental concepts to those of physics: the fact that propositional attitudes and related events and states are in part identified in terms of their causal and other relations to events extraneous in time and place... If externalisms of these sorts are... dominant and unavoidable features of the mental, the impossibility of incorporating psychology into a unified scientific theory of the world is clear.¹²⁶⁴

Davidson concludes that the “externalism” of mental states and events can be used to discredit type–type identity theories.¹²⁶⁵ This aspect of the irreducibility of the mental has been ignored by all those commentators who argue against the anomalism of the mental.

Davidson argues that propositional attitudes, the semantics of spoken words, and behavior, are all like the state of being sunburnt. The state of being sunburnt is necessarily caused by the sun, but it could be possible that two states of the skin could be intrinsically identical even though only one of them would be the state of sunburn.¹²⁶⁶ Davidson sees an interesting similarity between the state of sunburn and the state of belief when he notes that “my belief that moon is gibbous depend[s] in part on the causal history of my relations to the moon. But it could happen that two people were in relevantly similar physical states... and yet one could be... thinking of our moon, and the other not.”¹²⁶⁷ This view could be interpreted as an ontological claim, because Davidson speaks of states instead of concepts. However, as already mentioned, in the argument from externalism to irreducibility, two lines of thought, epistemological and ontological, can be detected. Let us consider the epistemological claim first. The main point of this claim is that in order to be able to identify mental states, an interpreter has to refer to the social and physical environments in which the interpretee has acquired these states. This is to repeat the claim that in belief attribution conditions of holism, externalism and normativity have an essential role, and there is no way to reject these principles as long as it is propositional attitudes that are being attributed.

This way of identifying beliefs does not have ontological implications; there is no reason to claim that the states so identified would not be subjective physical states which supervene on the states of brain. The causal history of mental states is relevant only in the sense that it affects the way in which we have to identify these states. Davidson notes: “[...] how we describe and identify events and states has nothing directly to do with where those states are.”¹²⁶⁸ Indeed, “[mental] states are ‘inner’ in the sense of being identical with states of

¹²⁶⁴ Davidson, 1997c, 71.
¹²⁶⁶ Davidson, 1995a.
¹²⁶⁷ Davidson, 1995, 122.
¹²⁶⁸ Davidson, 1988a, 52.
the body, and so identifiable without reference to objects or events outside the body...."\textsuperscript{1269}

The ontological claim, however, changes this picture. Davidson writes:

\begin{quote}
\ldots subjective states are not supervenient on the state of the brain or nervous system: two people may be in the same physical state and yet be in different psychological states. This does not mean that mental states are not supervenient on physical states, for there must be a physical difference somewhere if psychological states are different. The physical difference may not be in the person....\textsuperscript{1270}
\end{quote}

The reference is to subjective states which do not supervene on the states of the nervous system. Davidson is eager to emphasize that the physical states of two people on which mental states supervene cannot be \textit{identical}, because physical states cannot be identical unless they inhere in the same object. These physical states can be identical in \textit{“all relevant respects"}, and yet the psychological states may still be different. Davidson notes that this in fact is the position of AM. For two people to think alike, there do not have to be things – actual entities – which are or need to be identical. There cannot be identical physical entities, for the reason described above, and the expression \textit{“mental entity"} cannot be anything but a metaphor.

The quote clearly suggests that in Davidson’s view, the states of belief and other mental states are not supervenient on the physical states of an isolated agent. This in turn suggests that the \textit{individuation} of beliefs necessarily requires a reference to the outside world. There is an intrinsic physical state which \textit{is} the belief, but this state does not wholly determine what somebody believes. Whether your belief is the same as mine does not depend on our intrinsic physical properties, and therefore the question of whether our beliefs are the same cannot be resolved by considering only these properties. What is interesting is the claim that the mental states in question can be \textit{different} although the physical states of our brains are intrinsically identical. It seems that a specific mental state is identical to a physical state, even though it cannot be identified as mental without a reference to the reality outside the brain. This raises the ontological question of \textit{where} the mental states actually are. Are they wholly in the head or not?

What the ontological claim seems to make clear is that there is no hope of individuating mental states by referring to the intrinsic physical properties of an agent. But, this being the case, what does it mean to say that \textit{“[B]elief... is supervenient on facts of various sorts, behavioural, neurophysiological, biological and physical."}\textsuperscript{1271} On behavioural facts yes, but in what sense are beliefs supervenient, for example, on facts at the

\textsuperscript{1269} Davidson, 1987, 20, my emphasis.

\textsuperscript{1270} Davidson, 1989a, 62.

\textsuperscript{1271} Davidson, 1983, 147.
neurophysiological level? If beliefs cannot be individuated as neurophysiological events, in what sense are they then neurophysiological events? What is the reason for defending the supervenience relation between the mental and the neurophysiological? A belief, supposing that substance dualism is false, be supervenient on neurophysiological facts. But what about a belief of which an agent need not be aware? On what does such a belief supervene? Perhaps Davidson emphasizes the supervenience of beliefs on facts of various sorts merely in order to remind us of the truth of physicalism. If this aspect is left aside, I believe that the following is a better formulation of the nature of the supervenience relation between beliefs and “other levels of reality”: “[…] all there is for the… interpreter to get right... is what supervenes on causal interactions among speakers and the events and objects of their world.” Since the whole truth about the mental is captured by what the interpreter “gets” right, the whole truth about the mental rests on the supervenience relation between mental phenomena and behavior. Davidson claims that mental phenomena supervene on the physical properties of our bodies. It seems to me that this view is meant to emphasize the importance of the whole human being in the mind–brain debate. By arguing that our bodies constitute the “essential link between our minds and the rest of the nature”, Davidson places the behavioral, instead of the neural level in the forefront in a similar way to von Wright. This view stresses the primacy of interpretation and suggests what the main thesis of interpretative intentional realism should be: when we talk about beliefs let us forget the talk about neurophysiology and concentrate on the relations between speakers and the world. This view agrees with von Wright’s claim that the behavioral level has semantic priority with respect to the mental, which in turn has epistemic priority with respect to neural.

It seems to me that the ontological claim is indeed the argument that Davidson uses to defend the irreducibility of the mental. He says:

If mental properties are supervenient not only on the physical properties of the agent but in addition on the physical properties of the world outside the agent, there can be no hope of discovering laws that predict and explain behavior solely on the basis of intrinsic feature of the agents.

Many naturalists seem to think that the possibility of “scientific psychology” requires that there be correlations between beliefs and some intrinsic physical properties of agents. Chomsky, for example, suggests that a naturalistic study should concentrate only on the

---

1272 Davidson, 1999z, 81.
1273 See Davidson, 1993f, 295.
1274 Davidson, 1995, 122.
internal states of an organism. Fodor famously claimed that in cognitive psychology, the strategy of “methodological solipsism” should be adopted. Stich has argued that scientific psychology should ignore the semantic aspect of propositional attitudes. The examples are numerous. Davidson refers especially to Fodor and Chomsky when he says:

The reason thinkers like... Fodor and... Chomsky want to find a purely internal element or aspect of the propositional attitudes is obvious: it is only if mental properties are supervenient on the physical properties of the agent that there can be any hope of identifying the mental properties with physical properties, or of finding lawlike connections between the two.

The externalist feature of mental states makes their reduction to physical states problematic. I think that this is one of the most fascinating of the problems of reducibility, because it challenges the possibility that mental states could be reduced to the *states* of the brain. If I am correct about the ontological nature of this argument, it seems that it differs in important respects from the two previous arguments which base the irreducibility of the mental on conceptual reasons. Being a different kind of argument, the argument from externalism strengthens the overall case of the irreducibility of the mental. Perhaps the most interesting question about the possibility of reduction is whether we can literally point out within the brain those parts, entities or states which correspond to mental states. Neuroscientists claim that they are already doing this. But are they really pinpointing mental phenomena inside the brain?

If the externalism of mental states is true – if it is the case that beliefs do not supervene on the states of brain or nervous system – then it seems clear that beliefs will not be found inside the brain. This would dismiss a form of reduction which is based on the idea that all scientific truths should ultimately be explicable, in principle at least, by referring to fundamental laws governing the behavior of microphysical particles. If the ontological claim is correct, it seems that the microphysical particles on the basis of which scientific truths about beliefs could be expressed cannot be those of the brain alone. This would at least show that mental states are not “brain-reducible”. I will let the reader decide the importance of this view. The claim that beliefs do not supervene on the intrinsic features of a person connects with the claim made in section 3.2 about the priority of the mental over the physical. The beliefs simply cannot be found inside an isolated brain; the psychological level is required to verify the claim that a specific state of the brain is also a state of belief. If externalism is true, then the criterion of a belief cannot be found from the brain, just as von Wright claims.

\[^{1275}\text{Chomsky, 2000.}\]

\[^{1276}\text{Davidson, 1995, 122.}\]
One obvious consequence of Davidson’s commitment to externalism is that he has no use for “narrow content”, the search for which has been the task of many contemporary philosophers. States with narrow content would be needed for scientists to predict the behavior of an individual merely from the observation of the individual’s brain. The scientists in our laboratory example from section 3.2 would thus need the concept of a narrow psychological state. Davidson rejects the idea of a narrow content and thereby the possibility of any scientific psychology based on it because he does not understand what kind of states such narrow states could be. Davidson claims that these states cannot be beliefs or desires because the contents of these states are dependent on the causal histories of individuals, and each individual has a unique history. His conclusion is: “There are no propositional attitudes (the ‘narrow ones’) that are supervenient on the physical state of the individual taken in isolation from her history, society and environment.” Beliefs are not atomic features of the brain which could be considered individually in isolation from their propositional environments. The content of beliefs is not an item that can be defined in isolation from other beliefs. This is a straightforward consequence of interpretationism.

The rejection of narrow content has one interesting philosophical implication, namely the suggestion that a molecule-for-molecule duplicate of me would not necessarily have the mental states that I have. In some philosophical thought experiments we are sometimes asked to imagine that it is possible to create a perfect physical copy of Georg. If we are only physical objects, combinations of physical particles which function within the boundaries of physical laws, what would prevent our duplication? Could we not imagine that a perfect physical copy of Georg could be made if we knew enough of the behavior of physical particles and physical laws? Here the intuitions of philosophers differ about the mental states of the copy. Some, like Davidson, claim that since the copy of Georg lacks the kind of causal history that Georg has, the copy cannot have thoughts at all. A perfect physical copy would behave just as Georg does, and no one could tell the difference between them. Davidson’s conclusion is, however, that the copy would not have thoughts even if it seemed to have them. Many see this as an absurd conclusion, especially if the story is told in the way that makes it the case that the copy seems to have thoughts. From the perspective of the anomalism of the mental, the claim that physical identity does not entail mental identity is, however, a natural conclusion. In the context of externalism, this conclusion seems to be drawn from the view

---

1277 In Caorsi, 1999, 332.
that beliefs are essentially relational states, and copying the intrinsic properties of an object
does not necessarily mean that the extrinsic properties get copied too.

From the perspective of interpretationism, the conclusion is problematic however. It is
not clear how the claim fits with the position of “third-person Cartesianism”. If there is no
difference in the behavior of Georg and his copy and no one could tell, based on their
behavior, which one has beliefs and which one does not, what would it mean to say that
Georg actually has beliefs but the copy does not have them? I think that if we want to benefit
from these kinds of wild thought experiments, our conclusion should be robustly
metaphysical: contentful states without history are impossible. What makes a state have
content is the process of learning, in which certain states get to be connected to the
environment and thereby come to be about the environment. One may see this as an
implausible suggestion, but the alternative is to assume that states or objects with history
could be replaced with their ahistorical counterparts in a way that would make no difference
whatsoever. I don’t think this sounds plausible either. We are asked to imagine that an
intrinsic copy of anything physical is possible, in principle. Of course everyone agrees that
this kind of duplication will, most likely, never succeed. But why do we not want to say that
the duplication is not possible – is it because to become a certain kind of state requires a
certain kind of development through natural selection and learning? This would be nothing
more than a “metaphysical” interpretation of Davidson’s views about the sociality and
intersubjectivity of thought.

3.4 Conclusions of chapter three

In this chapter I have clarified the reasons for believing in the irreducibility of the mental. My
conclusion is that the arguments for the anomalism and irreducibility of the mental are not
conclusive. Davidson himself admitted that nothing in what he says about these matters can
be considered as a proof, and that there is no knock-down argument for AM. It seems
absolutely clear to me that the truth of the irreducibility of the mental cannot be proven. This
is true of most philosophical claims, and I believe that we should have an interest in the
arguments for irreducibility even though they do not conclusively show that the mental is
irreducible.

My reading of Davidson emphasizes that the irreducibility of the mental is based on
our choice to treat humans in a certain kind of way. If we want to see each other as rational
beings who can act out of reasons, we cannot get rid of our mental vocabulary. In my view the
“essence of Davidsonian irreducibility” is captured in the idea that our mental vocabulary is not an optional part of our conceptual resources. It is not a vocabulary which could be replaced with another, different kind of vocabulary without losing the explanatory power that the former vocabulary has. Davidson wanted to resist the irresistible ambition to consolidate our vocabularies of explanation, description and prediction. The important question is why this ambition should be resisted. The answer is that by getting rid of our mental vocabulary, we would leave behind something which we are capable of doing now. We can understand people in terms of mental concepts, in terms of their propositional attitudes. This kind of understanding would simply not be possible in terms of a language which did not refer to these attitudes and to the logical relations which exist between different mental phenomena. But it is just these relations which matter to our rational explanations, and rational explanations matter to us.

My interpretation is that all the reasons or arguments for the irreducibility of the mental serve the more general motivation to maintain our mental vocabulary. In section 3.1, we saw that the original argument for the anomalism of the mental in fact challenges only the view that psychophysical laws could be strict. Davidson’s claim is therefore much more principled and also much weaker than most interpreters have taken it to be. It is more principled in the sense that Davidson was interested in the question of whether there is some essential difference between the mental and the physical. As I showed in section 3.1, in Davidson’s view there is such a difference because physical vocabularies belong to the same conceptual domain, and therefore an in-principle reduction of physical vocabularies to the vocabulary of physics is possible. Our mental vocabulary will not reduce to physical vocabularies because of the disparate commitments of these different vocabularies. Again, our mental vocabulary will not reduce as long as we are not willing to change the subject, that is, as long as we are not willing to accept that mental phenomena could be given conclusive non-mental criteria.

The actual argument against psycho-physical laws described in sections 3.1.1 and 3.1.2 leaves something to be desired. It is definitely not a conclusive argument, but is based on the view that we cannot accept that our statements about the mental would be ultimately constrained by evidence from a different conceptual domain. As the discussion of section 3.1 showed, there are various confusions relating to questions of how the argument should be understood and how strongly it should be taken. I am not confident that I have been able to completely clarify the obscurity that most critics see around the argument for the anomalism of the mental.
The anomalism of the mental is, however, only one aspect of the non-reductive views of Davidson. In section 3.2, drawing on von Wright’s suggestion about the epistemic priority of the mental with respect to the neural, I clarified the claim that attributions of mental phenomena must be faithful to their respective sources of evidence. It was claimed that mental phenomena have behavioral criteria – and therefore facts about the brain, although serving as symptoms of the mental, cannot be the basic evidence on which attributions of the mental could be grounded. I think that von Wright’s contribution to the discussion concerning the irreducibility of the mental is extremely relevant. It shows something very basic about the mental–physical relation; from a logical point of view, the relation between mind and brain is accidental, whereas the relation between mind and behavior is conceptual and not empirical. This kind of philosophical insight is not appreciated by those naturalists who are interested in the mind–brain relation. In my opinion, von Wright’s claim that this is an empirical question and therefore not of interest for philosophy is a deep philosophical insight, the importance of which is not easily admitted by those who want to turn philosophical questions in scientific ones.

Davidson’s reasons for the irreducibility of the mental, which were discussed in sections 3.3.2, 3.3.3 and 3.3.4, can be seen merely as adding details to the view of why mental concepts cannot be eliminated or reduced to physical concepts. Davidson thought that holism, externalism and the normative features of the mental stand or fall together. These are elements of psychological concepts which cannot be eliminated without radically changing the subject. The dispositional character of mental concepts exhibits an important feature which cannot be eliminated from these concepts without losing their explanatory power. The normativity and rationality of thoughts enables us to see how their attribution is necessarily governed by the subjective norms of the interpreter. Seeing mental states from the perspective of the interpreter helps us to see how these states must be tied to the environment and to the history of the interpretee. This demolishes any hope of finding a purely subjective, narrow element of belief which would be implemented in the brain. The fact that the features of the mental stand or fall together means that one cannot easily reject the irreducibility of the mental without rejecting the whole interpretationist conception of the mental.

Many commentators claim that Davidson’s argument against psychophysical laws is successful only against traditional models of reduction, which require bridge-laws between the mental and the physical. This may be correct, but it is not clear whether it is a fair criticism against the Davidsonian position; it cannot be expected that an argument which was developed to refute a specific form of reduction could be extended to very different reductive
models. In sections 3.3.1.1 and 3.3.3.1 I showed, however, that two recent reductive positions are not entirely safe from the Davidsonian challenge. In order to see this, it is crucial to recognize that the argument for the irreducibility of the mental should not be equated with any specific technical argument defending such irreducibility. My claim is that the overall motivation to defend the view of why we need different kinds of vocabularies for explaining the nature of human life is Davidson’s most important insight. The reasons arising from this motivation are not taken seriously enough by those who would like to reduce minds to molecules.

In this chapter I have clarified the question “What are the reasons for the irreducibility of the mental?” and concluded that although a proof of irreducibility is too much to ask for, there are nevertheless severe problems facing the claim that the mental could be completely reduced to the physical. Mental phenomena and mental explanations are indispensable. This suggests that the truth of physicalism, which in fact does not get support from Davidson’s and von Wright’s arguments, seems to be even more unwarranted. The truth of physicalism cannot be shown. In the next chapter I will consider what kind of clarification the non-reductive position at which we have arrived can provide of the problem of mental causation.

Chapter four: The problem of mental causation and the possibility of epiphenomenalism

The main question of this chapter is: Can non-reductive physicalism solve the problem of mental causation? In the context of physicalism the problem of mental causation is clear. The problem is that of how phenomena intuitively taken to be mental can have causal efficacy in a thoroughly physical reality. In the context of physicalism, the general strategy for solving this question is also clear. We should find the neural correlates of mental phenomena and explain how these neural phenomena cause behavior. From the perspective of the problem of mental causation, all proposals about how to naturalize mental phenomena have the same purpose: They attempt to show how something physical can have the properties that we attribute to mental phenomena and explain how these mental–physical entities can have causal powers.
Recently there has been an increasing interest in the question of how mental causation can be understood in the context of physicalism. Kim is perhaps the most active philosopher who has discussed the problem of mental causation and its relation to physicalism. According to him, the problem of mental causation is this: “How can the mind exert its causal powers in a world that is fundamentally material?” For Descartes, who is often interpreted as the traditional source of the mental causation debate, the problem was how a non-physical mind can have an effect on the physical world. But in its modern form, the problem of mental causation is not really about how a non-physical mind can have causal powers. It is rather the question of how a physical mind can have causal powers. Kim refers to Davidson’s “Mental Events” as the source which reawakens the interest in the problem of mental causation, which had been a largely non-discussed issue for centuries. The problem is, however, that although AM established the physical nature of mental events, it left unanswered the question about the causal efficacy of these events. The current situation, owing much to “Mental Events”, is as Kim puts it, that: “...our basic physicalist commitments... can be seen as the source of our current difficulties.” Kim’s own view is that if a broadly physicalistic world view is accepted, then the solution to the problem of mental causation has to be reductive. This is a consequence of the more general naturalistic view that the only way to vindicate the status of mental phenomena is through reducing them to physical phenomena, by showing how the mind fits in the physical world. Kim concludes: “...if we want mental causation, we had better be prepared to swallow reductionism whether we like it or not.” The reason for this is that from the physicalistic principles which are required “for a physicalistic world view”, only one seems to be negotiable. Kim claims that the principles that cannot be rejected if a physicalistic world view is to be accepted are those of: a) physical causal closure, b) exclusion, and c) mind-body supervenience. The principle which has to be rejected is the one about mental–physical property dualism (to which non-reductive physicalism is committed). Non-reductive physicalism, if it accepts mental–physical property dualism, cannot adequately solve the problem of mental causation,

1280 Kim, 2001, 271.
1281 Kim, 2001, 272.
1282 Kim, 2001, 278.
1283 The principle of causal exclusion states that:” If an event e has a sufficient cause, c, at t, no event at t distinct from c can be a cause of e (unless this is a case of genuine causal overdetermination).” (Kim, 2001, 276)
1284 Although, as I argued in chapter two, it seems that many forms of non-reductive physicalism, for example the positions of von Wright and Davidson, are not vulnerable to the charge of property dualism.
because it is this very principle that has to be rejected to vindicate mental causation. I think Kim’s basic conclusion about the prospects of non-reductive physicalism is sound and well argued.\textsuperscript{1285} Non-reductive physicalism cannot provide a reductive solution to the problem of mental causation. This was something to be expected, because a reductive explanation would conflict with the most basic idea of non-reductive position. Kim’s claim is, as it were, a truism. But I think that the question of whether non-reductive physicalism should be rejected as a view about the mind because it cannot provide the kind of solution required by reductive physicalism is another one altogether. For the reasons to follow, we should disagree with Kim on his claim that non-reductive physicalism as a theory about the mind should be rejected. Kim’s main reason for rejecting non-reductive physicalism is that this position cannot explain how mental causation is possible.

A widely accepted contemporary view is that the solution to the problem of mental causation has to be \textit{reductive}. The conclusion is not surprising if we consider the reasons why reductionism is the suggested solution to the problem of mental causation. On the one hand, if we start with the view that non-mental reality is thoroughly physical (O-physicalism) and claim that only physical things can have an effect on the physical structure of reality (causal closure of the physical), then the conclusion is simple: mind \textit{has} to be physical \textit{if} it is to have causal efficacy in the physical world. Mental causation, far from creating a problem for physicalism, can be used as an argument which leads to physicalism – or at least to the mind–brain identity theory. This argument is based on the view that causal realm is physically closed. This, together with the view that mental causation is real, leads to a physicalistic conclusion. As we have seen, closure can be used to defend physicalistic intuitions in various ways. In chapter two it was shown how Davidson has been criticized based on the fact that if closure is assumed, monism follows. A similar kind of criticism can be applied here with respect to the problem of mental causation. If causal closure is assumed, a reductive solution to the problem of mental causation seems to be necessary. Mind is incorporated into physical reality through mind’s causal efficacy. The alleged efficacy of mind implies both that mental causation must be a form of physical causation and that mind, the seat of mental causation, must itself be a physical phenomenon.

On the other hand, if we start with the standard naturalistic view that mind is a physical thing, then a tempting and natural conclusion is that its causal powers are due to its

\textsuperscript{1285} Kim claims: ”…under nonreductive physicalism it is not possible to make sense of mental causation…” and mentions especially Davidson’s position. See the interview at http://www.ephilosopher.com/page.php?15. Kim (1995a) also rejects, for example, Searle’s biological naturalism because it cannot solve the problem.
physical nature. Why would mind be anything special in this respect if it is a physical thing? The complicated arguments, attempting to demonstrate that reduction is the only way to solve the problem of mental causation, seem unnecessary if either one of the above-mentioned physicalistic options are already chosen. If ontological physicalism about non-mental reality and closure are accepted, then the physical nature of mental causation follows. If the physicality of the mind is accepted, then too the physical nature of mental causation follows. In his latest book Kim summarizes his view by noting that his position: “... begins by embracing ontological physicalism.” But if this is so, the need to argue that mental causation needs “vindication” through reduction seems, if not unnecessary, at least irrelevant since the truth of ontological physicalism is already assumed. Mental causation is thus not needed to establish the truth of physicalism although it could be used for such purpose.

Kim admits that assuming the causal closure of the physical may sound question-begging to those who have a more dualistic view about the nature of reality, but claims that the argument for the need for reductive explanation is effective without this assumption as well. But it is nevertheless interesting that according to Kim the causal closure of physical reality comes back to this: “If a physical event has a cause that occurs at \( t \), it has a physical cause that occurs at \( t \).” It seems that this formulation leaves room for a mental cause that could also occur at time \( t \), and the formulation would thus not be question-begging from a dualistic perspective. But, in order to prevent causal overdetermination, Kim notes that we could adopt a stronger form of physical causal closure, stating that: “Any cause of a physical event is itself a physical event – that is, no nonphysical event can be a cause of a physical event.” A physicalist could simply adopt this stronger form of closure and thus straightforwardly reject the possibility of mental–physical causation. This would be wildly question-begging from a dualistic perspective. For this reason, Kim argues that “for a philosophical gain” the weaker form of closure should be chosen. This sounds strange. How could we choose the “strength of closure”? Isn’t the nature of this closure an empirical issue? As we remember, Kim has claimed that the closure principle is not an empirical issue for the physicalist. This means that a physicalist could as well accept a stronger form of closure. Is a weaker form of closure then accepted only “for the sake of argument”? So we could claim.

---

1287 “The principle of causal interaction” in Davidson’s argument is used in this way.
1288 It is clear that Kim himself accepts the closure principle so the problem of mental causation should not be a problem for him.
1289 Kim, 2005, 43.
1290 Kim, 2005, 50.
If a strong form of closure is rejected and a weaker form of closure is accepted, the argument for the need for reduction relies on the exclusion principle, which Kim in this context strangely describes as a “commonsensical assumption about causality.”1291 We are told that exclusion principle is: “...virtually an analytic truth with not much content....”1292 We should wonder what the argumentative value of such a truth is for the problem of mental causation, which is an empirical problem. Nevertheless, the argument, which combines a weak form of closure and the exclusion principle, has the consequence that we “know”, for example, that pain occurs only because a certain neural state occurs.1293 Suppose that pain causes your finger to twitch. According to Kim it is “highly likely” that there is a story about how a neural state caused the twitching of the finger and, by applying the exclusion argument, the only way in which pain can make a causal difference is by being the causally efficacious neural state. What follows from this, or so it seems to me, is a trivialization of the problem of mental causation. The combination of weak closure and the exclusion principle is no less question-begging than the argument which is based on a stronger form of exclusion. The (non-)problem of mental causation and its solution for the physicalist can actually be summarized by Kim’s words: “If mental phenomena are neural processes in the brain, there will be no special mystery about mental causation....”1294 This indeed seems to be the case if one accepts the if-clause – and accepting it requires no more and no less than accepting the truth of O-physicalism.

1291 Kim, 2005, 155. In another passage the exclusion principle is described as being a “general metaphysical principle” which does not favor physical causes over mental ones. (Kim, 2005, 17) It is hard to see that this principle would be a “commonsensical” assumption about causation because, quite likely, people usually do not have commonsensical assumptions about the nature of causation at all. A more serious problem is this. According to the exclusion principle if an event e has a sufficient cause c at time t, there cannot be an event distinct from c which could be the cause of e. But since the exclusion problem is neutral with respect to the mental and physical causes, why could not the excluding cause be mental? This possibility is denied only if ontological physicalism and the principle of causal closure are already taken for granted. 1292 Kim, 2005, 51.

1293 The possibility of a mental cause is thus excluded right from the beginning for reasons which have nothing to do with the exclusion principle itself. Now, do we really know that pain occurs because a certain neural state occurs and that pain without this state is impossible? I don’t know the answer but I cannot think of an argument which could show that pain has to be identical with a certain kind of neural state Kim (2005, 155) claims that we do not think that there could be sensations that could float free from the brain. This conviction, the acceptance of O-physicalism, makes the mind–body problem and the related problem of mental causation easy to answer. Even if the ontological claim is accepted there remains the conceptual point emphasized by von Wright; the criterion of pain is something else than a neural state. Therefore pain could – in some sense of the word “could” – be present in the absence of the neural state. What seems to be true at least is that, in principle, pain can be realized by different neural states and therefore it is not true that pain is identical with a certain kind of neural state. Therefore it is incorrect to say that “we know that pain occurs because a certain neural state occurs” because the expression “certain neural state” does not, in Kim’s jargon, refer to any neural state whatsoever, it refers to a specific state. Moreover, even if the ontological point is accepted with respect to sensations there remains the question whether the same point can be established in the case of other mental phenomena like beliefs.

1294 Kim, 2005, 153. The appeal of reductionism “…lies in its promise of a direct and simple account of mental causation.” (Kim, 1995a, 194, my emphasis) But perhaps it is a mistake to think that every problem should have a direct and simple solution.
It is easy to understand why a reductive solution to the problem of mental causation is the solution which non-eliminativist naturalistic philosophy has to aim for. A solution which did not naturalize mental causation by showing its proper place in the physical world would render the nature of this causation disturbingly mysterious and at odds with a scientific world view. I believe that since this says more about the commitments of naturalism than about the problem of mental causation, it is good to be aware of the healthy attitude that also is well described by Kim. Although he acknowledges that physicalism is the view in the philosophy of mind of the 20th century, he notes that “would-be physicalists” need to consider what kind of physicalism or how much physicalism it is possible to actually have. In this consideration, when the truth of physicalism is not obvious and not taken for granted, the problem of mental causation represents a profound challenge for physicalism. If the important physicalistic principle of causal closure or the truth of O-physicalism is not taken for granted, mental causation does not help the argument for physicalism. It is important to see that since the problem of mental causation is still unresolved, physicalism – if it is to survive as an all-encompassing world view – must explain where, if anywhere, we belong in the physical world (as Kim himself concludes). This requires showing, not merely assuming, that mental phenomena are in fact physically reducible. This is an empirical task. The problem of mental causation must be actually solved before ontological physicalism can be fully embraced. Therefore Kim’s own view, which starts by embracing physicalism, has the situation backwards. It expresses a commitment which is, but need not be, be taken for granted.

How about a position like Davidson’s or von Wright’s, which emphasizes the principled irreducibility of the mental, with respect to the problem of mental causation? I think there is only one answer: this kind of position simply cannot solve the problem of mental causation. Kim is right in his claim that mental causation creates insuperable difficulties for non-reductive physicalism. But, or so I claim, here we can and in fact have to make a choice. We can either reject the view that mental phenomena are irreducible or we can accept Davidson’s and von Wright’s arguments. Needless to say, these arguments are inconclusive, somewhat obscure and perhaps far from convincing. Because I admit that a non-

1295 It should be acknowledged that Kim’s (2005) last word on the prospects of physicalism is skeptical and he thinks that, pace many critics and commentators, the label “reductionist” does not well describe his position.
1296 We can thus say that the acceptance of the if-clause requires that one is able to show, not merely stipulate, that mental phenomena are physical phenomena.
1297 We can therefore absolutely agree that from the perspective of physicalism: “…anomalous monism falls short as an account of mental causation.” (Kim, 1998, 33)
reductive position cannot reductively solve the problem of mental causation, there is no need to participate in any discussion which focuses on the exclusion problem and related worries. From a Davidsonian–von Wrightian perspective this discussion is overtly metaphysical. I believe this is one important reason why von Wright and Davidson did not participate in the current mental causation debate. From their perspective, the discussion is built on certain confusions; the most severe of these confusions, so it seems, is the claim that we do not really understand the phenomenon of mental causation unless we give a reductive explanation of it.

Whether or not these arguments for the irreducibility of the mental are convincing, it is clear that this kind of solution to the problem of mental causation which, according to the naturalistic consensus should or must be searched for, is absolutely not an option for Davidson or von Wright. This being said, a curious caveat must nevertheless be noted. Kim claims that if mental phenomena are neural processes in the brain then there will be no special mystery about mental causation. If mental phenomena are neural processes then the ontological mystery of mental causation disappears. But according to Davidson and von Wright, mental phenomena are neural processes in the brain. By accepting token-identity they certainly agree with the view that mental phenomena are also neural phenomena. We have seen that the interpretationism of von Wright and Davidson goes against the view that mental phenomena are internal neural entities, but at the same time they seem to accept this identity in their monistic ontologies. If Kim thinks that the mental–neural identity solves the problem of mental causation then he should accept that the position of Davidson and von Wright; essentially the position of non-reductive physicalism avoids the mystery of mental causation in the same way that their views avoid the ontological mystery of the mind–body problem. But I think that a complete ontological solution to the problem of mental causation that would leave no puzzlement would require showing that mental phenomena are actually reducible to physical phenomena. This is not possible according to Davidson and von Wright. But neither is there any reductive approach in the philosophy of mind which has shown how to really reduce mental phenomena. From the ontological point of view, the positions of von Wright and Davidson are thus on the same level as reductive theories.

Because an essential aspect Davidson’s and von Wright’s views is the principled irreducibility of the mental, they cannot at the same time suggest that mental causation can be reductively explained or “vindicated”. Moreover, there is an important element in von Wright’s and Davidson’s views which shows the impossibility of a reductive explanation of mental causation but this, so I claim, is a positive contribution to the problem of mental causation. It is the claim that a reductive explanation of mental causation would not explain
what the relation of mental causation is meant to explain. The situations where we explain something by referring to mental causes could not be explained in a vocabulary lacking mental terms. Mental explanations explain phenomena in a different way than physical explanations. Mental causes – reasons – explain in a different way than physical causes; therefore a reduction of a reason to a cause would result in a loss of explanatory power. The conflict between irreducibility and the demand of naturalization is at the center of any discussion in which the merits of Davidson’s and von Wright’s views are considered. A position which aims at the naturalization of mental causation cannot tolerate the irreducibility they embrace, and takes non-reductive physicalism to be an unscientific view. Conversely, a position which argues for irreducibility sees the desire for naturalization as an expression of scientism. A reconciliation of these conflicting views seems unlikely. The starting points upon which the arguments rest are too different.

If it is argued that a mental–physical reduction is possible, the strategy for approaching the problem of mental causation is simple. The purpose is to understand how brain causes movements that we take to be voluntary and how mental phenomena that we subjectively experience relate to the brain events which can be objectively studied. If it is argued that a mental–physical reduction is possible, the strategy for approaching the problem of mental causation is simple. The purpose is to understand how brain causes movements that we take to be voluntary and how mental phenomena that we subjectively experience relate to the brain events which can be objectively studied. The focus will be increasingly inwards, towards the brain, and the entities under consideration are cells, neurons, and other microscopic features of the brain. EEG-curves, fMRI-scans and other results of neuroimaging techniques will be the evidence on the basis of which conclusions about mental causation are drawn. On this approach, mental causation is causation between physical entities, and because the purpose is to understand better the nature of physical causation, sciences such as chemistry and ultimately physics are relevant in addressing the question of how mental causation takes place in the brain. The benefits of this approach are clear; it promises to provide a coherent picture of our place in the universe. Human self-understanding increases when the prediction and control of our behavior becomes strict and when explanations become more accurate. The alleged promise of this approach is objective truth. A reductive explanation would settle the question of whether a mental cause could really be causally efficacious. The way to find out whether a mental cause can be really

1298 Fodor (1994a, 292), for example, notes that the agenda is to: “...Explain how minds qua material objects could have the properties they do” and continues in a naturalistic spirit that “...it is...reasonable to wonder whether this is a research agenda in philosophy.”

1299 Just to use one example. In “How Molecules Matter to Mental Computation” Thagard argues: “There is considerable evidence that chemical complexity really does matter to brain computation, including the role of proteins in intracellular computation, the operations of synapses and neurotransmitters, and the effects of neuromodulators such as hormones.” (Thagard, 2002, 429) The conclusion is: “[...] it may be time for...the philosophy of mind to become, like current biology and medicine, molecular.” (Thagard, 2002, 444) For a comprehensive attempt to turn philosophy of mind molecular, see Bickle, 2003.
causally efficacious is by considering whether the physical correlate of the mental phenomenon is causally effective.

I think that we should recognize also the risks of this program and of the underlying desire on which the program is based. Some of these risks are recognized by few of those who suggest that mind should be reduced to matter, but the possible concrete consequences of these risks are often neglected. Kim, for example, notes: “...reductionism, when applied to the mental, appears, at least to some, to have the unfortunate consequence of killing the patient in the process of curing him: in its attempt to explain mental causation, it all but banishes the very mentality it was out to save.” A non-reductive solution threatens to turn mind into an epiphenomenon without causal powers, whereas a reductive solution threatens to eliminate the mental altogether. When thinking about these “purely philosophical issues” we should recognize that the problem of mental causation is related also to broader questions. What if a reduction of mind to molecules suggests that most of the time we are not “in charge” of ourselves? Are we willing to accept this kind of situation, which may be an actual consequence of scientific research? Or are philosophers again needed to convince us that “compatibilism” is something which we can have faith in? It is not obvious that an explanation of mental causation in physical terms is something which should be pursued.

If mental phenomena are reduced to physical phenomena and if the physical nature of mental causation is shown, the possibility that we lose our freedom has to be considered. On the one hand, it is possible that through a reductive explanation of mental causation the possibility of objective explanation and control of mental causation is opened. Our sense of freedom, autonomy, and dignity would suffer if outsiders knew our inner life better than we know it ourselves, or if our actions were to become strictly predictable. On the other hand, finding the “true”, that is, physical mechanisms of human cognition may have the consequence that mental phenomena are left with no causal role to play. Perhaps we often misinterpret ourselves, our motives and our reasons for our actions. Sciences studying the brain could show that this is indeed the case. Neuroscience could show that the mental cause which we thought to be causally efficacious, and in terms of which we explained our action, was in fact a non-cause. This kind of result would certainly decrease our sense of freedom and could have a severe effect on our psychological well-being. It can also be imagined that describing us finally “as we truly are” may have the consequence that, in the name of consistency, our commonsense understanding of ourselves would have to be replaced by the

---

1300 Kim, 1995a, 194. The actual consequences of this conclusion are nevertheless not discussed further by Kim. Examples of these actual consequences will be considered in the appendix.
“true” description. A simple example of this could be a situation where a person sees a mental “defect” as an essential part of her personality while doctors or other specialists, knowing the physical cause of this defect, suggest that it should be “taken care of” or “cured.”\footnote{Consider Thagard’s suggestion that philosophy of mind should become molecular as medicine has already done.} The way in which mental illnesses are often treated in modern society is an example a situation where mental and physical facts are measured against each other and the former are seen as inferior in comparison to the latter. These and similar kind of risks are real, as the titles of recent books like \textit{The Illusion of Conscious Will}\footnote{Wegner, 2002.} or \textit{Living Without Free Will} show.\footnote{Pereboom, 2001.} These kinds of books show that philosophers and scientists are drawing drastic conclusions from pertinent empirical data. I believe the conclusions are such that they put human freedom into jeopardy. This is not terribly surprising, because the picture of reality which the natural sciences describe is mechanical and deterministic. When conscious free will, which exemplifies mental causation, is fitted into this picture a conflict is to be expected. The consequence is a human being which starts to increasingly resemble a physical mechanism – a machine.

One natural way to react to my criticism is to claim that scientific research must be allowed to take its course. If the progress of science shows that freedom is an illusion, then this is a conclusion with which we have to live. If freedom is an illusion it does not help to hope or wish that things should be otherwise. Enthusiasts claim that we must accept the consequences of scientific findings whatever they may be. But why should this be so? The question of whether scientific progress is always desirable requires a value judgment which cannot be decided through scientific research. But the need to make the choice based on values cannot be ignored; brain-research is already having concrete consequences which touch upon legal, social, medical and moral questions. If history is any guide to what will happen, it is plausible that these consequences will increase in the future and therefore a new kind of attitude is needed for new kinds of moral problems.\footnote{It may be thought that “objectivity” is what “scientific philosophers” search and that their work is therefore meant to be value-free. As we will see in the appendix, there are reasons to think that the opposite is true.} How we react to these consequences is an important question. We can appreciate also the view that it is important to clarify human nature in any way possible. Those who defend this view would perhaps claim that we have no choice but to stand with truth on this question. But we do have a choice. We can choose human freedom as a primitive concept and consider what the human condition is, or has to be, in a context where this basic fact is defended. It is not obvious that truth will set
us free, and those who want to follow all possible paths of scientific research must be willing
to take the consequences.

What kind of “form of life” we want to live is partly up to us. We do consider the
possible consequences of scientific progress – for example, the question of where limits of
gene-manipulation should be drawn or whether such limits should be rejected once and for
all. Why wouldn’t these considerations be relevant when it is consciousness, personality,
personhood which is at stake? The view that we are the controllers of mental causation is
essential for our self-image. In each individual case, it is the mysterious I who is in control.
Kim reminds: "[…] we care about mental causation… because, first and foremost, we care
about human agency."\textsuperscript{1305} In the end it is not the abstract demands of physicalism but the
importance of our self-image as free agents which leads us to contemplate the problem of
mental causation and which, in some philosophical circles, leads to attempts to “vindicate”
mental causation. I think it is this basic fact – human agency – which must be our starting
point, and all possible measures should to be taken in order to save it. Rejecting agency for
abstract metaphysical or methodological reasons is the last resort and even dualism is a better
option, \textit{pace} Kim’s fear. This kind of claim cannot, of course, be much further argued. But if
one is categorically opposed to this suggestion, it is because one has already accepted a
certain view about how things \textit{have to be}. But this view is not in harmony with the way we
think things are. We should remind ourselves of Putnam’s words: ‘The strength of the
‘Objectivist’ tradition is so strong that some philosophers will abandon the deepest intuitions
we have about ourselves-in-the-world, rather than ask (as Husserl and Wittgenstein did)
whether the whole picture is not a mistake.’\textsuperscript{1306} It is not the obscure desire to “save freedom”
that counts but actual human life, which shows that freedom is a non-negotiable fact about us
as long as we have, as we currently do, a certain view about ourselves. Our lives show what
we value and what we really believe. Some of these beliefs may be illusions, but a life with
illusions \textit{may} be better than a life without them.

A naturalistic philosophy of mind usually neglects these kinds of concerns
completely. Wittgenstein, Davidson and von Wright all expressed worries about the de-
humanization of people and emphasized that the possible harmful consequences of the
scientific image of man needed to be acknowledged. Wittgenstein took them \textit{very} seriously
and worried that philosophers are apt to miss what lies before everyone’s eyes. Of the
possible negative consequences of “objectivism” von Wright writes:

\textsuperscript{1305} Kim, 2002, 675.
\textsuperscript{1306} Putnam, 1987, 9.
That somebody else should have supreme authority in cases which concern my ‘inner life’ may be thought humiliating. May not such an authority misuse his position for ‘brain-washing’ – perhaps with a view to furthering uniformity in people’s thoughts and actions? And may not this lead to the gravest injustice in treating a person? Of these dangers we have good reason to be aware. 1307

According to von Wright there is something tragic about cases where a person’s own judgment about themselves is overridden by a “supreme authority”. What is so tragic about the situation? I think that one way to understand this is to claim that people should not be treated in this way; we should not think of each other as mere brains that can be manipulated. If person’s own judgment is constantly overridden, the result is likely a feeling of alienation and this is a direction which we should not take: it is not the way that people should be treated.

Although von Wright and Davidson raise worries against the consequences which may follow if our mental lives were reduced to nothing but brain-activity, this is not the reason why they cannot solve the problem of mental causation; my claim is that this is part of the reason. It is not the unwillingness to jeopardize human freedom which forces Davidson and von Wright to conclude that a reductive explanation of mental causation is impossible. This being said, it needs to be recognized that the reason for developing a theory like AM, in addition to securing the status of physicalism, was to secure the autonomy of agency. Davidson claims that: “Anomalism of the mental is… a necessary condition for viewing action as autonomous.”1308 He writes also: “…accounts of intentional behavior operate in a conceptual framework removed from the direct reach of a physical law by describing both cause and effect… as aspects of a portrait of a human agent.”1309 The fact that intentional behavior operates in a framework which, due to irreducibility, escapes the reach of physical laws, is what grants intentional behavior its autonomy. Von Wright, on the other hand, claims that: “It is not by being exempted from the bondage of natural law that man is a free agent.”1310 He emphasizes that freedom of agency is a result of our capability to understand man as a person, and we cannot understand the rest of reality and nature in this way. It could be tentatively claimed that we can understand man as a being with a “Wittgensteinian soul”, that is, our attitude towards other human beings is of a certain kind. When discussing the philosophical problems of mind, von Wright did not hesitate to occasionally use the term

1307 Von Wright, 1985, 26-27.
1308 Davidson, 1970, 225.
1309 Ibid.
1310 Von Wright, 1985 43.
“soul”. I believe this term, even when stripped of its religious connotations, captures something important of the attitude which we should have toward other human beings.

Davidson’s and von Wright’s arguments against reducibility have the consequence that a reductive explanation of mental causation is out of the question. The motives of Davidson and von Wright in defending irreducibility and thereby human freedom are of course interesting because, it seems to me, in the end it is the desire to defend freedom that leads to the arguments for irreducibility, which in turn leads to the rejection of reductive explanations of mental causation and thus to the securing of human freedom. The search for the motives of von Wright and Davidson need not concern us here. Here I merely acknowledge that they cannot accept a reductive explanation of mental causation because of their reasons for irreducibility. If we agree with Kim that mental causation should be reductively explained and that non-reductive physicalism cannot do this, we should ask whether positions like von Wright’s and Davidson’s can make any positive contributions toward “solving” the problem. I believe there are many such contributions.

The first thing that must be noted is that the nature of the problem of mental causation changes notably if seen from the Davidsonian–von Wrightian perspective. The new problem of mental causation, as I would like to call it, is the cluster of problems such as the following: What is the status of mental explanations? How should we understand these kinds of explanations? Can these explanations be trusted? Is there an essential difference between mental explanations which are made from the first- and third-person perspectives? There are many related questions that should be considered if the reality of mental causation is accepted and a reductive explanation is ruled out. I suggest that the most important question is: on what grounds can we believe that a mental explanation is true? Why should we be interested in this? Because we care about human agency; we want to know whether the mental cause in terms of which we explain our behavior can be trusted; we want to be transparent to ourselves. The question is important because it really matters to our lives. The question of whether or on what grounds I can trust my own self-explanations has profound importance for my psychological well being, and for my whole identity as a person who is driven by certain motives and desires. The question of how to trust mental explanations is the new problem of mental causation, and a reductive answer to it is not a possibility for a non-

---

1311 What such motives could be is nevertheless an interesting question. Why some philosophers end up defending eliminativism while others argue for non-reductive physicalism is a question worth considering. In the appendix I will briefly consider some possible motives of von Wright and Davidson.

1312 I use the expression “new problem of mental causation” to distinguish the problem from the traditional problem of mental causation which is the question how mental qua physical can cause.
reductive physicalist. Without mental–physical reductions the traditional problem of mental causation becomes a problem about the nature of mental explanations. On those occasions when we use mental explanations and think that they are true, we believe that we have detected a mental cause. This cause is whatever happens to be the mental entity in terms of which the explanation is given. This mental entity is a genuine mental cause, because its causal power cannot be explained physically. The causal efficacy of mental phenomena is grounded purely on mental explanations.

Those who reject Davidson’s and von Wright’s reasons for the irreducibility of the mental would claim that a reductive explanation of mental causation is possible. This would mean that the considerations about the new problem of mental causation are futile, because there is no such a problem. The truth of mental explanations is grounded on the fact that the mental entities to which the explanations refer have causally effective physical counterparts. Against this charge I reply that my interest in rejecting the traditional problem of mental causation and focusing on the new problem of mental causation is not based solely on the acceptance of von Wright’s and Davidson’s arguments. My claim is that even if a reductive explanation of mental causation could be possible, such an explanation should not be explored. Although I discuss the new problem of mental causation because it is a problem for Davidson and von Wright, the reason for considering this is more far-reaching. I will argue in the next sections that the new problem of mental causation and a form of epiphenomenalism related to it are something which should interest us because they have actual relevance for our lives. The question of how mental phenomena qua physical phenomena can be causes is instead solely a philosopher’s problem; it is an important problem but it does not have relevance for our daily lives. Already Descartes noted that the problem of mental causation becomes a problem for us only when we start to philosophize. An updated version of this observation could be the view, inspired by Wittgenstein, that mental causation becomes a problem for us only when we start to philosophize in an unhealthy way. The new problem of mental causation is instead a deep problem which may affect the way we live.

The perspective that I am suggesting could be called an epistemological or deflationary approach to mental causation. That the problem of mental causation should be carried on in terms of epistemological considerations instead of metaphysical ones is a strategy defended for example by Burge and Baker. My interpretation is that von

---

1313 For this label see Walter, 2007.
Wright’s approach to the problem of mental causation belongs to these epistemological or deflationary approaches. His claim that the mental has epistemic priority over the neural has the consequence that the assumed causal connection between neural processes and muscular movements cannot be the warrant of truth of the rational explanation of an action. We can recall Davidson’s suggestion that, in the case of mental properties, we can “[…] concentrate on what might be called the epistemological problem and let the ontological problem, if there is one, take care of itself. For I think that if we were to solve the epistemological problem we would lose interest in the supposed ontological problem.”1316 I admit that in the context of O-physicalism metaphysical questions are part of the problem of mental causation. There are metaphysical aspects of the problem just like there are metaphysical aspects of the mind–body problem which cannot be completely rejected merely by accepting the position of interpretationism. Some of these problems are challenges for a non-empirical position like von Wright’s and Davidson’s if these positions contain views which compete with empirical claims. But I also believe that by considering the epistemological question we are able to see that mental causation does not need vindication through reduction. Burge suggests that we:

[…] take as our philosophical starting-point, not a metaphysical doctrine about the nature of causation or of reality, but a range of explanations that have been found worthy of acceptance. They… include commonplace explanations that explain the phenomena that we encounter in everyday life…. If we put aside the metaphysical picture and begin with the explanations that work, causation becomes an explanatory concept.1317

Baker has a similar view:

If we put aside the metaphysical picture and begin with the explanations that work, causation becomes an explanatory concept…. If we reverse the priority of explanation and causation that is favored by the metaphysician, the problem of mental causation just melts away.1318

This is a profoundly important insight. The overtly metaphysical approach to the problem of mental causation, which has survived thanks to the domination of naturalism, is a modern result of philosophers’ tendency to imitate science (which, in Wittgenstein’s words, is the real source of metaphysics). The demand for reductions is an unhealthy sign. We can and should consider the problem of mental causation without being biased by a metaphysical framework of materialism; the reasons for believing that even O-physicalism is true are unconvincing.

1316 Davidson, 1995b, 44.
1317 Burge, 1993, 92-93.
1318 Baker, 1993, 93.
If we are convinced that reduction of the mental to the physical is impossible, the way to approach the new problem of mental causation is certainly not simple. The purpose is to understand how a mind–body system, an entity called person, functions in a mental–physical reality. A person, a *psychophysical unity*, is not clearly separable into “mind” and “body”. A conclusion resembling this kind of view was reached when discussing the nature of Davidson’s and von Wright’s ontological positions. The final consequence of taking interpretationism seriously is that the problem of mental causation becomes the problem of how people can trust the mental explanations in terms of which they explain themselves and others.

4.1 What happens to mental causation if the mental is not an ontological category?

Before I present a view which emphasizes the epistemology of mental causation over the metaphysics of mental causation, I will consider Davidson’s “official solution to the traditional problem of mental causation. He, unlike von Wright, did participate in the contemporary debate, which focuses on the metaphysical *problematique* of mental causation. Davidson did this in order to defend AM against critics who argue that AM cannot solve the problem of mental causation. The more severe and very common complaint is that AM implies the causal inertness of mental phenomena. But how should we think of the problem of mental causation from a Davidsonian perspective if the irreducibility of the mental is accepted? What can be said about this problem if the mental is seen as a conceptual and not an ontological category?

Different questions comprise the problem of mental causation. On the one hand it can be asked how mental events, in virtue of being mental, can cause something. On the other hand, the modern challenge seems to be the question of how mental phenomena, in virtue of being physical, can cause something. This is the question that especially worries naturalists. At least three different stances towards the problem of mental causation are possible. Perhaps they do not exhaust the alternatives but, rejecting substance dualism, it is difficult to think of a proposal that would not fall, in one way or another, among the following suggestions.

1) The question of how mental phenomena cause things cannot be separated from the question of how we give mental explanations. Causal efficacy of mental phenomena is tied to the way in which we explain or understand something as
being performed because of mental causes. Causation and causal explanation
cannot be clearly separated.

2) Mental phenomena cause things because they are identical to physical phenomena
and physical phenomena have causal powers.

3) Mental phenomena do not cause anything. They are epiphenomenal.

Although I believe that an epistemological or deflationary approach to the problem of mental
causation should be defended, it still seems to me that each of these suggestions contains part
of the possible truth. In the following I will clarify this claim and show how difficult it is to
conclude what Davidson’s view about mental causation is. Those critics who see AM as a
version of epiphenomenalism have oversimplified both Davidson’s views and the problem of
mental causation.

The three solutions above can be all true at the same time. Let us consider a case
where it seems to me that I do something because I believe B and desire D. I go to a museum
because I believe there is currently a Picasso exhibition and I want to see his paintings.
According to 1), if I go to the museum the explanation that I give of the situation warrants my
saying that my belief–desire pair caused my action. The reason that I believe that B and D
cause my action is just the fact that I can give an explanation in which I refer to B and D and
I take this explanation to be true. If I could not offer this kind of explanation we could not
even start to ask whether B and D were causal factors in my behavior, and the question of
mental causation would not arise. In the 1960s the famous debate between Davidson and the
intentionalists began from the question of how to distinguish efficacious reasons from
existing reasons. Davidson’s view is usually interpreted as a suggestion that an underlying
causal relation secures the truth of reason-explanations. I believe, however, that we should
read Davidson in a different way. As in the case of von Wright, the truth of rational
explanations is based on the facts of interpretation. It could be said, as Davidson is interpreted
as claiming, that if B and D really caused an action, there must be an underlying causal
relation between the action, B and D. But this does not explain at all why or how we take
rational explanations to be true. The expression “really” in this context is meaningless.
Stoutland’s perspicuous analysis has convinced me that Davidson’s position has been widely
misinterpreted.1319 I believe that both Davidson and von Wright want to defend the view that
the truth of rational explanations does not depend on the assumed causal connection between
reason and action. Our acceptance of the truth of the claim that I went to the museum because

I wanted to see the Picasso exhibition does not depend on our conviction that a causal relation exists between the reason and the behavior. Of course, there are cases when I have reasons to go the museum and I go there but not because of these reasons, and there are cases when I have reasons to go the museum and I do go there because of the reasons. The causal theory does not explain how we could really distinguish these cases.

But the claim that a mental explanation is all we need to draw the conclusion that B and D caused my behavior does not exclude the possibility that B and D caused my behavior, because they are physical states which can be causally efficacious because of their physical properties. Answer 1) does not exclude answer 2), and if we want to know how B and D caused my action it may be the case that the answer must refer to physical mechanisms which are currently unknown. But whatever the future of mind–brain identity turns out to be, our understanding about the nature of reason explanations and our belief in them does not depend on it. In “Actions, Reasons and Causes” Davidson famously claimed that the laws which back up rational explanations use concepts of neurology, chemistry or physics. He insisted:

Nothing I have said implies that we cannot give causal scientific explanations of particular human actions, thoughts and the like… if we choose a particular event, there is no reason why we cannot say this particular event (and here we describe it with a physical description) is a causal factor in the production of this mental event… there is no reason… to suppose that we cannot name the causes of mental phenomena.

Davidson’s later position is suspicious of this view. I think that the mental became more irreducible than the previous quote suggests. On the other hand, the above view does not say much about the possibility of mental causation. It could be easily admitted that a physical event is a “causal factor” in the production of a certain mental event; this is what we do all the time. Davidson’s suggestion that a causal scientific explanation of a particular human action or thought could be given is certainly more problematic. If we ask how my belief B caused my other belief C or my action A, and require that the answer must be given in physical terms, it is difficult to think what kind of answer would satisfy our interest. I admit that I cannot imagine what kind of answer would be satisfying.

It is possible to defend the idea that the mental is causally efficacious in virtue of being identical with the physical while insisting that, because of the irreducibility of the

---

1320 Davidson, 1963, 17. A similar point is repeated when Davidson (1999, 654) notes: “[…] either can we infer from the fact that what we typically think of as the reasons for an action are not events that there aren’t events that are essential to the causing of the action. These events may or may not be mental in any ordinary sense.”

1321 Davidson, 1964, 48.
mental, we can never say how mental qua physical causes things. This is a very disappointing result for those who would like to describe how the mental aspects of an event are related to the event’s physical aspects. This ontological disappointment can be lessened by noting that our inability to give this kind of description does not mean that there is no such a relation. The fact that we cannot describe how mental qua physical causes does not mean that mental causation is an illusion if causation is understood in terms of explanation. The proponent of 1) could conclude that questions about “causation itself” cannot sensibly be asked outside the context of causal explanations. Mental explanations warrant the talk of mental causation but strictly speaking there is no such thing as mental causation.

A proponent of 2) can accept all of this and contend that those events which can be described in the mental vocabulary have causal efficacy because they are physical events and, by definition, physical events have causal efficacy. What he would perhaps deny is the claim that the truth of mental explanations does not depend on the underlying physical mechanisms. In his opinion it is precisely the working of this mechanism which warrants the legitimacy of mental explanations; there is no real mental cause if it is not backed up by a causal relation between physical events. As a reply, the proponent of 1) could ask whether the value or relevance of the explanation “I went to the museum because I wanted to see a painting of Picasso” would be lost if it turned out that I had, let’s say, no nervous system at all. The answer to this question is not obvious; if we explained person’s behavior in terms of his beliefs and desires and it turned out that he has no brain, would we say that all the explanations were false? It certainly depends on our concept of causation. Wittgenstein, for example thought the most important thing that needed to be done in explanations of thought and brain-processes was to reject all the old prejudices about causality. He famously claimed that it is perhaps time for our concept of causality to get upset. Perhaps we are not willing to take this drastic step. The question of how we would react if we were to find out that a certain person has no brain at all or that his brain was made of silicon is nevertheless interesting. Would we conclude that all mental explanations are false? If so, should we now conclude that all our current mental explanations may be false because we cannot be sure that other humans have brains? If we were to discover that the brains of others were made of silicon or plastic, would our puzzlement be less than in the case where nothing inside the skull could be found? A mechanistic view about causation requires the postulation of a mediating mechanism, but is this mechanism necessary for us to believe in mental explanations?

According to proposal 3), mental phenomena do not cause anything. This claim can be understood in different ways. One way to make sense of it is to say that the mental aspects of
a phenomenon are irrelevant to what the phenomenon causes. In the example considered here, this could mean that the content of my belief, namely “that there is a Picasso exhibition in the museum” is irrelevant to what I do. A defender of this view could claim that my behavior, a trip to the museum, has a physical cause but the cause is not related to the fact that I believe something. A relaxed form of epiphenomenalism could be defended by claiming that usually our beliefs do not cause anything, but sometimes they do, and when they do it is because they are physical things. Yet another approach with an epiphenomenalist streak is to claim that we have absolutely no way of telling when our beliefs cause something and when they don’t. This possibility being the case, the threat of epiphenomenalism, or the threat that our beliefs are irrelevant in each specific case, is always a possibility. This is compatible with the view that we can have well-functioning mental explanations, the adequacy of which depend on the connections between the concepts used in these explanations.

My claim is that, surprisingly, Davidson can be interpreted as holding each of the views 1–3. First, what is Davidson’s relation to the view that the problem of mental causation cannot be separated from the ways in which we give mental explanations? He says, for example, that: “We can specify the logical relations between the propositional contents of the appropriate beliefs and desires and the description of the action under which they rationalize it. What we apparently can’t do is say in clear detail how the mental attitudes cause the action.”

One of Davidson’s central claims in the 1960s was that reasons are causes and that reason-based explanations are a form of causal explanations. In true reason-based explanations, belief and desire cause the action. But our knowledge of the correctness of a mental explanation, or our belief in its truth, does not depend on our knowledge of how belief and desire cause the action. Indeed: “[…] we can explain behaviour without having to know too much about how it was caused” and this is the main point of reason-based explanations. This connects with the claim that mental concepts are dispositional concepts; explanations in terms of them are not full in the same sense as the explanations of physics are. There is no way to check, outside the context of mental explanation, whether the belief and desire really cause the action to be explained. This was something vigorously emphasized by von Wright, who noted that in reason-based explanations the causal hypothesis rests on the acceptance of the reason-based explanation and therefore, quite obviously, the warrant of

---

1323 Davidson, 1974b, 232.
1324 To state the issue in these terms (does x really cause something) is misleading because this question does not make sense if the view that the causal efficacy of the mental is tied to mental explanations is accepted. When explanation instead of causation becomes the primary concept, the “real-question” does not make sense.
truth of the reason-based explanation cannot rest on the causal hypothesis. If the official Davidsonian line is taken, then it can be claimed that if a belief causes an action then there is a causal relation between them. But it is not possible to tell, independently of a reason-based explanation, whether a belief causes an action, and therefore whether there really is a causal relation between mental phenomena and the action in question. This means that the question of mental causation is in the end a question about the nature of reason-based explanations. This in turn means that mental causation should be insulated from physiology; the phenomenon of mental causation should be freed from the grip of a picture which is a result of causation is seen in the physical sciences. The attempts of naturalists to analyze mental causation come to a dead-end when faced with the question: Did belief B really cause action A? An attempt to answer this will refer to some reason-based explanation and it will have to take this explanation as its starting point. This is von Wright’s point about the epistemic priority of the mental as applied to the problem of mental causation. I gladly admit that my way to state the issue may ignore the philosophical motivation which guides certain philosophers’ attempts to analyze mental causation. I nevertheless believe that the motivation itself should be challenged; the problem of mental causation should be considered outside the jungle of empirical metaphysics by taking into consideration the primacy of our actual practices – as Baker and Burge among others have concluded. We should not let abstract metaphysics be our guide because, as Burge notes: “[…] the probity of mentalistic causal explanation is deeper than the metaphysical considerations that call it into question.”

What is Davidson’s relation to the second view, which claims that the causal efficacy of the mental is secured because mental phenomena are physical phenomena? Here we again see the relevance of the question of what Davidson’s form of monism actually is. If physical monism is accepted then the claim that mental phenomena can cause and do cause in virtue of being physical phenomena is understandable. Indeed, the truth of physicalistic monism is important because “the ontological reduction, if it succeeds, is enough to answer many puzzles about the relation between the mind and the body, and to explain the possibility of autonomous action in a world of causality.” The interpretation that Davidson defends is justified given that he says:

[...] supervenience implies that if two events differ in their psychological properties, they differ in their physical properties (which we assume to be causally efficacious). If supervenience holds, psychological

1325 See section 3.2.
properties make a difference to the causal relations of an event, for they matter to the physical properties, and the physical properties matter to causal relations.\textsuperscript{1328}

This sounds like a claim that events have causal relations because they have physical properties; it is at least the way that most commentators have read Davidson. It is difficult to say how to exactly understand the expression “matters” in this context. One plausible way to understand the claim is to conclude that causal relations obtain \textit{because} of physical properties. For example, if we want to say why a piece of chalk broke when it hit the floor, we would have to refer to the \textit{microscopic properties} of the chalk.\textsuperscript{1329} This is a plausible claim given a broadly physicalistic view of the world. Why would the situation be any different in the case of mental phenomena if mental phenomena just are physical phenomena? The claim that causal relations obtain because of physical properties is also an easy answer to the question of why AM is not consistent with epiphenomenalism. It is because “[a]nomalous monism holds that all mental events, at least all those that enter into causal relations, are identical with physical events. Therefore they are caught up in the exactly same causal nexus in which we assume physical events are caught up.”\textsuperscript{1330} There is only one structure of causal relations and events exist in this structure because they are physical. This is not to say that we could not understand events as being causes also when described in the mental vocabulary. Given what we know about Davidson’s physicalism, the following view is understandable: “Anomalous monism makes sense of the claim that attitudes are dispositions to behave in certain ways, which are in turn physiological states, which finally are physical states….\textsuperscript{1331} Given this view, there is no reason to doubt that a physical mechanism could explain \textit{how} something mental caused what it did.

What is AM’s relation to epiphenomenalism? Davidson is not defending epiphenomenalism but he makes statements which could be interpreted as pointing towards a form of epiphenomenalism. Although Davidson argues that AM is not consistent with epiphenomenalism he also admits, surprisingly, that: “[Anomalous monism]… is consistent with the (epiphenomenalist) view that the mental properties of events make no difference to causal relations.”\textsuperscript{1332} However, AM does not \textit{imply} epiphenomenalism because Davidson

\textsuperscript{1328} Davidson, 1993d, 197, my emphasis.
\textsuperscript{1329} See Davidson 1964.
\textsuperscript{1330} In Caorsi, 1999, 338. Consider also: “Cause is the cement of the universe; the concept of cause is what holds together our picture of the universe, a picture that would otherwise disintegrate into a diptych of the mental and the physical” (Davidson, 2001, xv) If epiphenomenalism were true, the mental would float away from the reach of the physical resulting in a diptych of two substances.
\textsuperscript{1331} Davidson, 1997c, 72.
\textsuperscript{1332} Davidson, 1993d, 196.
thinks that 2) is correct. But the situation is not so straightforward because Davidson also notes that: “There is a harmless sense in which one can say the contents of the reasons on which we act is irrelevant to the causation of the action, because we can, in theory, explain why the action occurred using only physical laws and descriptions.”

My interpretation is that here Davidson departs from von Wright, who wants to make a firm distinction between reasons and causes as well as between behavior and action. From von Wright’s point of view, Davidson is here guilty of conceptual confusion when he says that an action could be explained using only physical descriptions. It is of course possible that by the term “action” Davidson is referring to the bodily movement; if this is the case the term is ill-chosen in this context. Likewise, a comment that: “Someone can describe an act of mine… in terms of movements described behavioristically, or in the language of physics and so on” is obscure. Perhaps the claim should be read as a suggestion that the event which can be described as an action can also be described in other vocabularies. But, as far as I understand, an action could not be described as an action in a non-mental vocabulary. In order to describe the event which is my action in the vocabulary of physics, one would first nevertheless have to identify the action in the familiar way; this I take to be the crucial lesson of von Wright.

Davidson nevertheless claims that the contents of the reasons on which we act is irrelevant to causation. Those, like Dretske or Fodor, who claim that an argument against epiphenomenalism requires precisely an explanation of how the contents of reasons are relevant would surely be dissatisfied with Davidson’s position. The use of the term causation is problematic because Davidson’s claim really seems to be that the fact that my mental state has the content which it has is irrelevant to what it causes. Perhaps it is really the case that Davidson does not see the mental as being efficacious, at least not qua mental. Even though we have not seen a coherent view which would explicate what this qua mental could mean in the context of O-physicalism, perhaps the claim that the content of a belief is irrelevant to what it causes suffices as a view which deprives the mental aspects of an event of their causal powers. What could also be considered as a statement of epiphenomenalism from Davidson’s part is the following claim: “Since beliefs, desires, and intentions aren’t entities, it is a metaphor to speak of their changing and hence an extension of that metaphor to speak of them as causes and effects.”

The claim that the talk of beliefs and desires as causes and effects is an extended metaphor sounds puzzling. It seems to me that there is a kind of constructivist

---

1333 In Caorsi, 1999, 338, my emphasis.
1334 In Caorsi, 1999, 335.
1335 Kalderon, 1987, claims that Davidson’s accounts fails to explain the causal efficacy of semantic content.
1336 Davidson, 1999, 654.
element in Davidson’s view about mental causation. He notes that we cannot talk about mental causation if the cause is thought of in a “neutral mode” so that the cause is disregarded its mental status as belief or other mental phenomenon. In order not to think of the cause in a neutral mode, “we introduce a mental description of the cause, which thus makes it a candidate for being a reason.”1337 This of course goes well together with the claim that events are mental only as described, or with the view that in radical interpretation behavior is taken as data on the basis of which a set of attitudes is hypothesized. These hypothesized attitudes help the interpreter to explain what he has observed. The claim that the set of attitudes of the interpretee is always a hypothesis could be likened to von Wright’s important observation that the attributions of mental phenomena are always accompanied with a residue of meaning. It seems to me that, on a Davidsonian–von Wrightian account, from a third-person point of view the attitudes which explain actions are hypothetical entities. As Davidson concludes: “We, watching a creature adjust its behavior to its needs and opportunities, read into that behavior the beliefs and desires and intentions that naturally occur in animals with thoughts.”1338 Here we should recall the discussion of section 3.3.2 where the subjective aspect of mental state attributions was emphasized. What we read into the behavior of others is a result of our subjective considerations about how to best make sense of the subject in a particular situation. Whereas Davidson’s views here are consistent with interpretationism, it is not so clear what these views tell us about the reality of mental causation.

This brief description of Davidson’s relation to 1–3 should be enough to show that, given his physicalistic monism, he can accept all of the three solutions to the problem of mental causation and can accept them together without being grossly inconsistent. Unfortunately the situation is not so straightforward because in addition to 1–3 there is Davidson’s “official proposal”.

4) Events are causally efficacious. The causal relation is an extensional relation between events and its nature cannot be further analyzed.

I believe this view of Davidson is best understood in relation to his version of neutral monism. Causal relations obtain between events which are non-abstract particulars, but there is no ultimate explanation of why or how they hold. The problem is that one can argue in favor of 1—3, where the second view expresses a commitment to physicalistic monism, or

1337 Davidson, 1982a, 180.
1338 Davidson, 1999c, 207.
one can argue in favor of 1), 3), and 4) where the fourth view expresses a commitment to neutral monism, but one cannot consistently defend all solutions 1–4. So, either 2) or 4) must be rejected. Which one should Davidson reject? Which one should we reject?

What exactly is Davidson’s official position? Two of its most important claims are that events are non-abstract particulars and that extensional causal relations exist between them. In “Thinking Causes”, Davidson emphasizes these claims because he thinks that an ignorance of them is the main reason which has lead critics to accuse AM from epiphenomenalism. Given these two views Davidson is entitled to say: “[…] given my concepts of events and of causality, it makes no sense to speak of an event being a cause ‘as’ anything at all.”

The fact that an event is a physical event cannot explain everything that the event causes and this goes against view 2). It makes no sense to say that an event caused something “as physical” or “as mental” Indeed: “It is events that have causes and effects. Given this extensionalist view of causal relations, it makes no literal sense… to speak of an event causing something as mental, or by virtue of its mental properties, or as described in one way or another.”

Davidson’s general claim is that it makes no sense to say that an event causes anything in virtue or its properties or that it causes something because of its properties. This, according to many, is an astonishing view. As Welshon comments:

I find this proposal incredible… although… events themselves have causal powers and are causally efficacious, nothing mental or physical about them is that in virtue of which they have those causal powers, and nothing mental or physical about them is that in virtue of which they are causally efficacious. But on the face of it, this claim seems fantastic, for, if true, [an]… event’s causal powers would be something brute about it, inexplicable by any science, including physics.

Welshon makes an interesting comment when he notes that, on Davidson’s account, the causal powers of an event would be inexplicable even by physics. This cannot be entirely correct because Davidson claims that physics could cover reality in full. But, as discussed in section 2.1.3 Davidson’s example of a strict law does not use causal concepts. It merely asserts that certain states follow from certain others. In a sense this is not a causal explanation at all, because it does not answer the question of why state B followed state A.

How could Welshon’s puzzlement be clarified? Given Davidson’s “nominalism”, his view is that properties cannot cause anything because they are abstract objects. Individual things with those properties can, however, be causally efficacious. The rejection of the causal

---

1339 Davidson, 1993d, 188.
1340 Davidson, 1993d, 196.
efficacy of properties may seem harsh, but I think there is a closely related insight which is interesting. It is this: does it make sense to ask, in the case of a particular unrepeatable event, what its causally efficacious property was, i.e. what was the property that was causally efficacious? As Davidson asked in “Causal Relations”: “How could Smith’s actual fall, with Smith weighing, as he did, twelve stone, be any more efficacious in killing him than Smith’s actual fall?” How could we say that the property of weighing twelve stone was the causally efficacious property responsible for Smith’s death following his fall? It was the fall which caused the death. Given that Smith’s fall was an event of a man weighing twelve stone falling, shouldn’t we rather say that: “[…] every property of every event is causally efficacious”, as Davidson does? Because Davidson thinks that every property of an event is essential to it, in his view it makes no sense to choose properties which could be somehow more causally efficacious than others. A critic like Welshon claims that if this road is taken, then AM is saved from epiphenomenalism but only at the cost of accepting that events do not cause other events because of their properties. The claim that an event’s causal powers would be something brute about it is a “fantastic and incredible suggestion” according to Welshon, but others have recognized that this is indeed a view that Davidson seems to defend. Child, for example, notes that the relation of causation can be taken to be a basic, natural relation between events which does not hold in virtue of anything else. He compares it to other natural relations like temporality. When $a$ precedes $b$, the temporal relation between them does not hold in virtue of anything else more basic; its holding is itself a basic fact. Marras notes that for Davidson, the causal relation is: ”[…] a fundamental, primitive relation, not one to be explained or analyzed in terms of more basic nomic relations between properties.” Commentators who ignore this are simply accusing AM from a perspective which a proponent of the position would not accept.

Davidson never tried to analyze the causal relation very clearly, perhaps because it appears on his list of concepts which cannot be reduced to more basic concepts. Let us ignore here the question of why the relation of causality cannot be analyzed. For our purposes it suffices to note that Davidson does not analyze this relation, and therefore critics who argue that, according to Davidson, events can enter into causal relations only in virtue of, the fact that, or because, their strictly nomic properties are mistaken. This applies, for example, to

---

1342 Davidson, 1967a, 150.
1343 Davidson, 1987a, 113.
1344 Marras, 1999, 276.
1345 It has to be noted that certain views of Davidson certainly justify this interpretation.
the critical views of, Sosa, Stoutland, Johnston and Kim. As Marras observes, AM in itself: “[…] has nothing to say about the causal powers of properties, mental or physical, or about what it is ‘in virtue’ of which events have their causal powers.” This being the case, the views from “Thinking Causes”, which emphasize that the relation of causality is extensional, certainly make sense. Davidson claims that, given his view of properties, he has never argued that events are causes because of their physical properties.

Campbell has taken Davidson’s later position seriously and recently defended AM as a non-reductive solution to the problem of mental causation. His position is interesting and, so it seems to me, a quite accurate reading of Davidson’s views – but the view is problematic for one reason. The reason is that Campbell attributes to Davidson a strong form of nominalism. Although there are reasons to claim that this attribution makes sense, there are also reasons, as I showed in section 2.3.4, to think that Davidson’s view of properties is not as straightforward as Campbell takes it to be. Campbell rejects the standard criticism against AM on the grounds that it does not take Davidson’s nominalism seriously enough, and thereby attributes to Davidson a metaphysical position which he does not hold. Whereas this is a fair comment, it is not clear whether nominalism can serve as a starting point for a viable account of mental causation.

If Davidson’s nominalism indeed commits him to a version of neutral monism in which causality holds between events as a brute fact, the remaining question is: is this the final word or does the physicalistic interpretation make sense after all? To answer this question we have to try to separate the questions of causal explanation and causation.

4.1.1 Davidson on causal explanation and causation

According to my interpretation, at least as long as we stay inside the Davidsonian perspective, the questions about causation and causal explanation cannot be clearly separated. Davidson claims that if an event A causes an event B, then this is true no matter how the events are described. This is turn suggests, or could be taken to suggest, that it is something in the events which is responsible of the fact that the one causes the other. But since it is in a sense up to us

---

1346 See Sosa, 1984, Stoutland, 1985, Johnston, 1985, Kim, 1993. I claimed earlier that, according to Davidson, “causal relations are physical relations.” Since I am here only trying to explicate Davidson’s position (4), I am not taking back this claim. Indeed, as I noted, only one of 2) or 4) can be true and the other must be rejected. My claim is that we should perhaps accept 2). The view that causal relations are physical relations agrees with this position.

1347 Marras, 1999, 277.

1348 Campbell, 2008.
what we count as an event, it seems that events themselves do not have clear spatio-temporal boundaries. This being the case, we can say that once we have sliced up reality into certain events, it is then something inside those slices which explains why one slice causes another.

Let us suppose that we want to claim that an event A causes another because of its physical properties and that those properties are intrinsic to the event in the sense that: a) they are inside the spatio-temporal slice that we count as an event A, and b) those properties exist regardless of us. Let us accept for the sake of argument that this is a rough description of how we think about the nature of causation or causal relations. But what does it mean to say that the properties of the event are out there regardless of us, if we choose what counts as an event on the occasion to be explained? Let us imagine that we have identified an event – an avalanche – which we take to be the cause of another event – the destruction of a hut. Which properties of the avalanche are causally efficacious on this occasion? Davidson would say that if the avalanche (event A) caused the destruction of the hut (event B), then this is true regardless of how we happen to name the events. But what does this claim tell us about the question of whether A caused B? If A causes B, there is a relation holding between them – but do we have any way to tell when or whether there really is such a relation? Is our only way to get a grip on that relation through causal explanation? Is the existence of the relation derivative from the fact that we think that A caused B on certain occasion? If so, what reason is there to insist that the causal relation is an extensional relation?

Let us imagine another example: every day when I look out of my window at noon I observe the following: a man walks down the street and when he passes a certain store, the door of the store opens. Is there a causal relation between these two events? Perhaps, but how am I to tell? I have observed these two events occurring every day for ten years and this could be a reason for me to conclude that the one causes the other. But it is of course possible that their simultaneous existence is an accident and there is no causal relation between the events. As Hume pointed out a long time ago, the causal relation cannot be observed. I believe the question of how to find out whether two events are causally related becomes especially pressing in the case of mental causation. Of course, in the case of the man and the store door, I could say that if the appearance of the man causes the store’s door to open then there indeed is an “extensional relation of causality” between the events, but does this claim say anything beyond the triviality that “A causes B iff A causes B”? Davidson claims:

[…] we must distinguish firmly between causes and features we hit on describing them and hence between the question whether a statement says truly that one event caused another and the further
question whether the events are characterized in such a way that we can deduce… that the relation was causal.1349

But what are causes without the features in terms of which we describe them? When can we say truly that an event caused another? It seems that we cannot in any clear sense speak of a cause without identifying, through using a feature of the cause, something as “the” or “a” cause. This being the case, it seems that the concept of a causal relation sounds like an abstraction, and nothing can be said of it without a causal explanation.1350 When Davidson says that “Causality relates events however described…”1351 or that “Causal relations… have nothing to do with how we describe or classify events…”1352 what can this possibly mean? What are these relations? Why do they hold? These claims could be interpreted along the lines proposed above in a) and b), but since Davidson does not recognize the existence of properties it is difficult to think that he would allow properties to be something in virtue of which events stand in causal relation to each other. If there are no such things as properties does it make sense to ask which properties are causally efficacious and which are not? No. But if it is not in virtue of causally efficacious mind-independent properties that events cause, then what does it mean to speak of a causal relation which is independent of our ways of describing and classifying events? Davidson claims that: “[…] if causal relations are ‘in nature’, it makes no sense to classify them as logical or contingent”.1353 In my opinion this claim includes a problematic proviso; if. The question of whether causal relations which hold independently of us exist, is similar to the question of whether it makes sense to speak of events outside all descriptions. It is clear that somehow Davidson tries to maintain this idea of neutral events, but it is difficult to understand what they are because events are always described as mental or physical. As I showed in section 2.6, formulating the nature of neutral monism is not easy.

There are therefore two deep problems for Davidson’s position. On the one hand, there is the question of whether anything about causal relations can be said outside the context of causal explanation. If not, what is the point of saying that causal relations are extensional while causal explanations are intensional? On the other hand, the question is why – i.e., in virtue of what – or how are two events causally related? Suppose that we want to explain a

---

1349 Davidson, 1967a, 155.
1350 I noted that Davidson does not want to analyze the relation of causality. But why then say that it is an extensional relation? Since we lack a detailed analysis of the relation, what would be a reason to believe that this claim about extensionality is true?
1351 Davidson, 1993e, 312.
1352 Davidson, 1993e, 313.
1353 Davidson, 1985, 224.
specific event, say the death of Georg on 1.1.2000. We know that he died from a heart attack. But consider the many ways that we could start to try to explain his death. I already gave a good explanation; the cause of death was a heart attack. This explanation satisfies our epistemic interests. Most of us understand this explanation, but it is not the most detailed explanation available. We could ask, when told that Georg died of a heart attack: “Yes, but why?” We have identified a particular event, which is Georg’s death. Should we say that the cause of the death was a heart attack or that the heart attack was identical with the death of Georg? We should say that the heart attack was the cause of Georg’s death and not identical with it, because the former occurred before the latter. But then we could ask how the heart attack caused the death, or what caused the heart attack. Both questions are, or may be, relevant when explaining Georg’s death. If we focus on the second question, the question about our explanatory interests becomes especially relevant. We ask what caused the heart attack of Georg’s, which is a particular event existing at a particular time in a particular place and occurring for a period of time.\footnote{What caused the heart attack is obviously a different question than what causes a heart attack.} In considering this question we can make the sphere of explanation larger or smaller. Larger, if we start to consider, for example, why Georg’s heart was in the kind of condition it actually was. Smaller, if we start to consider, for example, how a heart attack is related to the functioning of blood vessels and cells. If we expand the explanation we could be interested to hear that Georg refused to obey his doctor’s orders, that preceding the heart attack Georg had a stressful weekend or that his father died of a heart attack at young age, and so on. If we want to reduce the sphere of explanation we could be interested to hear that there was a blockage in a certain artery, and that this blockage was a result of such and such chemical processes, which in turn were a result of such and such lower level processes and so on.

If we are now being asked why Georg died or what was the cause of his death what are we to say? Should we say that he died because there was a blockage in his artery or because such and such chemical processes happened, or because he had a father with certain kind of genes, or because he refused to listen to his doctor, or because he had a very stressful weekend? In some sense the claim that Georg died because of a heart attack is our best explanation; we can say that the heart attack was surely a cause of the death. But was it the cause? In the case of a unique unrepeatable event, which every event is, does it make sense to speak of the cause? If we have a reason to believe that Georg’s refusal to obey his doctor was a causal factor in his heart attack, it seems that we are entitled to say that this refusal two
years earlier was, partly, what caused his eventual death. How we choose to answer the question “Why did Georg die?” seems to depend, partly, on what we take to be an adequate or a good answer. In explaining something we may choose the cause according to our special explanatory interests. In some context, a reference to Georg’s refusal to obey his doctor may be a better answer than a reference to the condition of Georg’s vessels. If we believe that ignoring the advice of the doctor was a causal factor in Georg’s death, it would be strange to say that the disobedience did not cause his death because the death was caused by a blockage in his artery. Almost anything can stand in the relation of causal explanation to something. My claim is that the relevance of a certain explanation, and therefore the relevance of a certain causal relation, depends on whether or not it improves our epistemic situation. Whether it does so, can be evaluated only subjectively.

I think there is something attractive about the idea that events do not stand in causal relations to each other because of their properties, or better, that if they do, it is because of all of their properties, as in the case of Georg’s death. Should we say that all the factors to which we can refer in explanation were part of the cause? If we consider any actual event occurring in the real world, it is clear that the event under consideration is unique. Events in time can happen only once. The dramatic idea behind this view becomes obvious if we think, for example, of meaningful moments of our lives, the uniqueness of which are clear to us. If we consider an event which is the first meeting of two people we can ask: why did they meet, why did this event occur? What was the cause of this event? It is possible to become anxious when it is realized that the meeting was, in a sense, the sum of endless and arbitrary events. When the situation is considered it becomes clear that there is no single answer to the question: ‘what was the cause of a specific meeting at a specific time in a specific place?’ If nothing can be pinned down as the cause, we could just as well say that everything preceding that specific moment in the history of the universe was relevant for the fact that the moment occurred.

---

1355 A good question is whether this should be thought of as a cause or as the cause, and whether there is a meaningful distinction to be made between the two in this context.

1356 Many would be willing to analyze this question in terms of counterfactuals, but it seems hopeless. Had the bus broken down he would not have been in place X at time t. Why assume that this analysis would be correct in a reality in which the bus had broken down? The nature of reality would have been different; perhaps the man would have taken a train or taxi instead.

1357 Although it seems clear that everything cannot be relevant for the fact that a specific meeting of two people occurs. The meeting takes place at 1.1.2000. How could a small meteor hitting the surface of Mars on 1.1.1970 be relevant for the meeting? The relevant boundaries must be drawn somewhere, but how are they to be drawn? On the other hand, perhaps the meteor, through a freakish causal chain, had some relevance. Can we completely exclude the possibility that this is not the case when it comes to a unique, specific event?
Every event is *unrepeatable*. This clarifies Davidson’s claim that every property of every event is essential to it, and that every property of every event is causally efficacious. From this view it follows that: “If we consider an *event* that is a “full, sufficient” cause of another event, it must, as Mill pointed out long ago, include everything in the universe preceding the effect that has a causal bearing on it, some cross section of the entire preceding light cone….“1358 A man is killed with a gun. Intuition perhaps tells us that had the gun been equipped with a silencer, the death would have nevertheless occurred if the silenced shot were similar to the original shot in other respects. Given that every property of an event is essential to it, Davidson concludes that in this case the second shot would have resulted in a death, but it could not have been the *same* death as caused by the loud shot – nor would the shots have been the same.1359 If this is true then it really does not make sense to say that the loudness of the shot was irrelevant to what the shot caused, because the shot that was fired was in fact silent and it caused what it did. Had it been loud it would not have caused the *same* things.

So, is it possible to answer the question “What is it about events c and e that makes it the case that c is the cause of e?”1360 *from Davidson’s perspective?* What was it, after all, about the avalanche that caused the destruction of the hut? What was it about the heart attack that caused Georg’s death? I think that we have to conclude that this question does not make clear sense from the Davidsonian perspective. But if this is the case how should we interpret Davidson’s claim that we assume that physical properties are causally efficacious and that physical properties matter to causal relations? Well, Davidson argues that: “Properties are causally efficacious if they make a difference to what *individual* events cause….“1361 The problem with this view, together with the view that the laws which connect causally related events deal with “microscopic properties” of the objects involved in the events, is the issue of how well it fits with Davidson’s dismissal of properties in favor of predicates. It surely sounds strange to say that “*Predicates* are causally efficacious if….“ I think that the only way to understand the claim that physical properties are causally efficacious or matter to causal relations is to say that *physical predicates* are “causally efficacious” if they matter to *physical explanations* and that causal relations cannot be understood a being anything else than *explanatory relations*. A causally efficacious property is nothing but an aspect of an event which we take to be causally efficacious, and this aspect we choose according to our

---

1358 Davidson, 1993d, 199. Consider also, the cause of a specific action: “The “cause” must embrace everything in the universe within the sphere defined by the distance light travels in the interval from $t$ to $t'$ from the region of the cause to the region of action.” (Davidson, 1999p, 639-640)

1359 The example is Sosa’s. Davidson discusses it in 1993d.

1360 Kim, 1993, 22.

1361 Davidson, 1993d, 198.
explanatory interests. Causally efficacious properties are thus completely explanation-dependent. The questions of what kind of events exist, where the boundaries of events are to be drawn, and how many events occur on a certain occasion are up to us. We have to conclude that there is no such thing as causation separated from our explanatory practices, and there is no point in talking of causal relations as separated from these practices. Crane has complained that Davidson’s position: “[…] seems to leave us unable to answer the question of why certain explanations are better than others by invoking the efficacious features of reality.” But what features could be invoked independently of explanatory interests? My claim is that certain explanations are better than others because they satisfy our epistemic curiosity in a better way.

We can now re-analyze the Davidsonian solutions to the problem of mental causation. According to my interpretation the options were the following:

1) The question of how mental phenomena cause things cannot be separated from the ways in which mental explanations are given. The efficacy of the mental is tied to the way that we explain or understand something as being performed because of mental causes. Causation and causal explanation cannot be clearly separated.

2) Mental phenomena cause things because they are identical to physical phenomena and physical phenomena have causal powers.

3) Mental phenomena do not cause anything. They are epiphenomenal.

4) Events are causally efficacious. The causal relation is an extensional relation between events, and its nature cannot be further analyzed. There is no answer to the problem of mental causation beyond the answer that events can cause things.

In our re-analysis, let us for the moment ignore the third possibility. This is obviously one answer to the problem of mental causation, but it is not a satisfying answer for those who believe that mental causation is real. From what has been said above, it should be fairly clear that proposal 1) seems to be by far the most plausible.

According to proposal 2), mental phenomena can cause things because they are identical with physical phenomena and physical phenomena are causally efficacious or, as Davidson reminds us, because physical properties are causally efficacious. The relation of supervenience is meant to connect the mental properties of an event with the event’s physical properties.
properties. But, as we have already seen, Davidson’s view of properties makes this kind of answer difficult to understand. In fact, whereas the standard formulation of supervenience in the contemporary philosophy of mind is made in terms of physical and mental properties, Davidson, when being precise, formulates his version of supervenience in terms of predicates. He writes: “[…] a predicate \( p \) is supervenient on a set of predicates \( S \) if and only if \( p \) does not distinguish any entities that cannot be distinguished by \( S \)”\(^{1363}\). This is one important way in which Davidson has expressed his view about supervenience. As the discussion of sections 2.1.2 and 2.3.6 showed, sometimes Davidson speaks of supervening properties or “features”\(^{1364}\). He notes that a feature \( F \) may supervene on a set of features \( F_1 \) even if \( F \) cannot be defined in \( F_1 \) or related nomologically to features in \( F_1 \). According to Davidson this describes the nature of the mental–physical relation. The relation could be called anomalous supervenience, as Davidson himself does call it.\(^{1365}\) I am not aware that anyone else other than Davidson would have used this expression. Anomalous supervenience goes well together with AM, of course; it is unclear, however, how this understanding of supervenience fits with the more usual understandings. This may be yet one more issue where commentators misinterpret Davidson’s position.\(^{1366}\) On the other hand, “anomalous supervenience” may give credence to Hare’s complaint, mentioned earlier, that Davidson managed to turn a very clear notion into a mysterious one.

Let us assume that the formulation of supervenience in terms of predicates instead of properties or features is the best way to characterize Davidson’s understanding of supervenience.\(^{1367}\) A consequence of this is that it is perhaps impossible to compare Davidson’s view to the views which take a more metaphysical attitude towards supervenience. This kind of comparison cannot lead to constructive results if both sides hold on to their respective views about the nature of supervenience and their underlying metaphysics. This being the case, I believe that some of the criticisms against Davidson’s position are irrelevant, except from a perspective which here turns out to be question-begging. It is disappointing that critics see Davidson’s position only from their own perspectives, which in turn makes the criticism insignificant as many commentators have noted.\(^{1368}\) The criticism is, of course, appropriate if it is meant to question whether Davidson is right – but

\(^{1363}\) Davidson, 1993d, 187.

\(^{1364}\) Consider: “Mental concepts are supervenient on physical concepts in this sense: if events fail to share a mental property they will fail to share at least one physical property” (Davidson, 1995c, 207).

\(^{1365}\) In Caorsi, 1999.


\(^{1367}\) “The notion of supervenience, as I have used it, is best thought of as a relation between a predicate and a set of predicates in a language”, as Davidson (1985, 242) notes.

\(^{1368}\) For an insightful discussion see van Brakel, 1999.
usually the criticism is based on a straightforward assumption that his view about properties is incorrect. This leads to a situation where critics fail to appreciate the interesting aspects of Davidson’s view and conclude that his proposal is “incredible”. Van Brakel has concluded, when referring to the dispute between Davidson and his critics about mental causation that: “[t]he critics, after twenty years of studying…” ’Mental Events’, don’t seem to understand [Davidson’s position] and don’t really address it.” I believe this is a fair statement; I have showed how Bickle and others do not really address the issues that Davidson sees as preventing the reducibility of the mental. Those who accept a different view of properties than Davidson does do not usually give reasons for this. It is not clear that Davidson’s alleged skepticism towards properties is less justified than a straightforward acceptance of properties.

It is not surprising that the primacy of interpretation, which is a central theme in Davidson’s philosophy, plays an essential part also in the question about the efficacy of the mental. Some commentators have noticed this. Campbell points out that Davidson’s version of supervenience: “[…] is not metaphysical dependence but is rather a form of semantic dependence since the relation is one between mental and physical predicates.” Mental predicates are attributed according to the general principles of interpretation. These attributions are based on physical evidence; changes in physical facts concerning, for example, the behavior of the interpreter or his environment will affect what mental predicates can be attributed to him in a way that respects the principles of interpretation. There is a very general relation of dependency between the mental and the physical, but this is a different kind of relation of dependence than suggested by those who have a metaphysical view of supervenience and who claim that mental properties supervene on physical properties. Davidson notes that: “[…] an object cannot alter in some mental respect without altering in some physical respect.” But as I argued in 3.3.4, the physical change which is required for a mental change does not need to be in the object. Objects may differ mentally while being physically identical, but if there is a mental difference or change, there must be some physical difference between the objects. Davidson notes that there must be a physical difference somewhere, but the difference may not be in the person. The difference may be, for example in the personal histories.

We are justified in attributing mental states to an interpreter on the basis of her behavior, and behavior requires physical changes. Without these changes we have no

---

1371 Davidson, 1970, 214.
1372 Davidson, 1989a.
evidence, or to use von Wright’s term criteria, for mental attributions. Therefore the mental predicates that can be applied to an interpretee depend on the physical predicates that can be applied to her. In my view the semantic dependence, which in Campbell’s view describes correctly the nature of Davidsonian supervenience, could be seen as describing also von Wright’s view. Although von Wright, when discussing the nature of supervenience, used the expression “supervenient in time”, I fail to see why the semantic relation between the behavioral and mental which is essential to von Wright’s position could not be understood in terms of supervenience. Davidson’s anomalous supervenience is not a reductive relation and neither is the semantic relation – with its residue of meaning – between the behavioral and the mental. If we recall how von Wright stresses the semantic connection between the mental and the behavioral, we should conclude that the views of Davidson and von Wright about the nature of supervenience are similar. This is at least one way that von Wright’s position could be understood, and related to the contemporary discussions about the nature of supervenience and the problem of mental causation. Von Wright’s conclusion is that: “since the relation of behaviour to the mental is semantic and not causal, we cannot by transitivity conclude that the neural is cause of the mental, nor, needless to say, the mental of the neural.”

It is not clear that the general dependency of mental on the physical should be seen as giving any priority to the latter. Campbell claims: “Since interpretation must begin with physical evidence, I take this dependence of the mental on the physical to be a fundamental feature of Davidson’s theory.” But as I showed in section 2.5.2 when discussing the nature of radical interpretation, it is not obvious that the evidence for the attribution of mental predicates is entirely physical. The data in radical interpretation is partly physical, partly mental and perhaps the best way to describe the relation between the mental and the physical is along the lines that von Wright suggests, namely as being a conceptual mixture. But it is clear that Davidson’s understanding of supervenience differs importantly from the usual understanding. Van Brakel concludes:

If any particular supervenience relation is relative to a passing theory of interpretation… and if there is little support to assume there are strict laws, this provides a very different perspective from the mainstream concern with supervenience, which is wedded to a micoreductionistic Scientific Image.

There are good reasons to agree with van Brakel and Campbell given Davidson’s views about the nature of the mental. I must add, though, that sometimes Davidson himself seems to see

his version of supervenience as a more metaphysical relation than it is, thereby giving support to the views of his critics. This becomes obvious from Davidson’s admittance that his version is very close to Kim’s notion of “weak” supervenience which is, after all, a thoroughly ontological view. Kim has affirmed that he sees Davidsonian supervenience in this way. It is no wonder that Kim can therefore raise criticisms against this notion, because Davidson’s version of supervenience does not agree with Kim’s intuitions. Whereas Kim claims that supervenience is best understood as a reductive relation, we have seen that Davidson uses the term “anomalous” when describing his understanding of supervenience.

Davidson’s use of the terms of his critics is unfortunate and causes confusions. It also leaves the reader under the impression that Davidson himself was confused about his own position. As Klagge has noted:

\[\ldots\] there are two very different notions going under the name ‘supervenience’. There is [a] methodological constraint on a theorizer – and there is Kim’s notion – a substantive constraint on the relationship of facts in possible worlds. Davidson’s view’s on supervenience uneasily straddle these different notions.\[1377\]

Davidsonian supervenience could be understood, interestingly, as a methodological constraint. On the other hand we could conclude, as de Caro has perspicuously done, that in Davidson’s writings two different concepts of supervenience can be detected. But what if supervenience is thought of as a semantic relation? In von Wright’s case this is the route that is taken. Can something else be concluded from the suggestion that Davidson’s supervenience or von Wright’s mental–behavioral connection is a semantic relation?

That supervenience is best understood as a semantic relation is another reason to claim that it is not the causal efficacy of physical properties which has a role in Davidson’s account of mental causation. Supervenience connects predicates, not properties. This supports the view that when we talk of causation we are usually talking of causal explanation, because predicates are relevant for explanations. Against my interpretation that causal relations are actually explanatory relations, Davidson would say that: “causal relations… are not in any

\[1376\] An example of this kind of confusion is expressed by Kim. He says: “Davidson is unlikely to feel comfortable with unrestrained talk of properties; he would perhaps prefer to talk of predicates instead.... Davidson interpreters may need to be careful about such issues. However, we will carry on our discussion in terms of properties and other reified entities.” (Kim, 2003, 135) But if the issue is more substantial, then it is not just “Davidson interpreters” who have to worry about it, because then the criticism based on certain understanding of the nature of supervenience simply does not apply to Davidson’s position. Campbell correctly notes that: “Since Davidson’s version of supervenience relates predicates rather than properties, and the sort of dependence this expresses is semantic rather than metaphysical, Davidson need not be troubled by Kim’s worries about modality.” (Campbell, 1998, 38)

\[1377\] Klagge, 1990, 347.
direct sense language-dependent”\textsuperscript{1378} and that they are somehow “in nature”. But what does this mean if causal relations do not hold in virtue of the physical and mind-independent properties of events? We can tentatively accept Davidson’s view that causal relations are “in nature”, but there is no answer to the question of why there exist such relations, and there is nothing that could be specified in virtue of which the relation holds. For these reasons, it seems to me, saying that these relations are “in nature” is pointless. This being the case, the answer to the problem of mental causation cannot really be proposal number two, which ties the causal efficacy of the mental to the causal efficacy of the physical. Strictly speaking, an event’s causally efficacious physical properties are inferred from explanations into which the descriptions of other relevant events enter because they satisfy certain predicates. The properties referred to by the explanatorily relevant predicates are then treated as being causally efficacious.

Proposal 4) seems also to be an unsatisfying answer to the problem of mental causation. According to this proposal, mental events can cause something because it is in the nature of events to cause things. Therefore it should be the case that the fact that an event is mental or physical is irrelevant to its causal efficacy. But in Davidson’s view, supervenience ensures that if two events differ in their psychological properties they also differ in their physical properties and, since the latter are causally efficacious, it seems to be the case that mental properties matter, after all, to the event’s causal efficacy. But this suggestion cannot work if proposal 2) is rejected. On the one hand, Davidson tries to say that only events have causal efficacy but, on the other hand, they would not have the causal powers which they have if they were not the events they are, and they would not be the events that they in fact are if they lacked some of the physical or mental properties which they actually have. It is difficult to see how these views could be reconciled. How does the view that “[…] if causal relations and causal powers inhere in particular events… then the properties we happen to employ to pick them out or characterize them, cannot affect what they cause”\textsuperscript{1379} fit with the claims that “…properties are causally efficacious if they make a difference to what individual events cause, and supervenience insures that mental properties do make a difference to what mental events cause”\textsuperscript{1380} or “Either they [mental properties] make a difference [to the causal relations of an event] or they don’t; if supervenience is true, they do”\textsuperscript{1381} Since Davidson wants to formulate his thesis of supervenience in terms of predicates instead of properties, it

\textsuperscript{1378} Davidson, 1985, 224.  
\textsuperscript{1379} Davidson, 1993d, 190.  
\textsuperscript{1380} Davidson, 1993d, 198.  
\textsuperscript{1381} Davidson, 1993d, 197.
makes no sense to say that *mental properties* do make a difference to what *mental events cause*.

We could accept proposal 4) as a suggestion about why mental events cause things, the answer being that it is because they are events, but this is not a clarifying answer. It would be almost like saying that mental phenomena cause things just because they do. If the claims that events have the powers to cause things is meant to be an answer to the problem of mental causation, we are back to the issue of how we explain or understand certain events as being causes of others. When we take events to be the relata in causal relations, we are once again in the realm of causal explanation because when we start to talk about events we have to talk them under certain descriptions, and which descriptions we choose depends on our explanatory interests. Putting things this way once again raises the question of what Davidson’s view about the nature of events is. Are they physical in themselves or are they neutral? If the mode of being of events were physical then we could perhaps defend proposal 2), which claims that events cause things because of their physical properties. This proposal is rejected by Davidson, and therefore the interpretation that events are neutral gains more credibility. But the neutral status of events is a formulation which remains obscure, because whenever one starts to talk about them, one identifies them under their physical or mental descriptions.

### 4.1.2 Davidson on mental causation and rational explanation

The conclusion of the previous section was that the best answer to the problem of mental causation that can be found from *Davidson’s views* was the proposal that causation cannot be separated from explanation. The question of how mental events cause cannot be answered outside the context of rational explanations. “Mental *qua* mental” causation is to be answered in terms of rational explanations, whereas “mental *qua* physical” causation is a confusion when seen from a non-reductive point of view. Causation and causal explanation cannot be clearly separated, and the expression ‘mental *causation*’ becomes obscure. As Davidson reminds us: “[…] it makes no literal sense… to speak of an event causing something as mental…”\textsuperscript{1382} Yet, he uses expressions like “[…] mental properties do make a difference to what mental events cause”\textsuperscript{1383} or “… self-criticism… is clearly a case of mental causality that

\textsuperscript{1382} Davidson, 1993d, 196.  
\textsuperscript{1383} Davidson, 1993d, 198.
transcends reason…”¹³⁸⁴ and “… certainly the ways metaphor, imagination, conceptual creativity and daydreaming work their wonders in the mind are cases of mental causality….”¹³⁸⁵ Mental cause is defined by Davidson as a state with propositional content, and he claims that in standard reason-based explanations the propositional contents must have appropriate logical relations to each other. The states of belief and desire must however also cause the explained state or event.¹³⁸⁶ Although there is the demand that a belief or other mental phenomenon must cause the explained event, it can also be said that reasoning is a process which can be thought of as entirely mental, as Davidson notes.¹³⁸⁷

But how should we understand the expressions mental causality or mental cause if it makes no sense to speak of an event that causes something as mental? Isn’t mental causation exactly an example of an event causing something because it is a mental event? The only way to make sense of this apparent problem is to note that the word “cause” can be understood in different ways. Davidson notes that: “Everybody allows that most talk about causality is interest-relative; what we call “the” cause of some event is some feature chosen from the totality of causal factors which particularly interests us….”¹³⁸⁸ This observation we made in the case of Georg’s death. We choose the cause of death from a point of view of some particular interest.¹³⁸⁹ Some critics suspect that if an event has all its properties essentially and if every property is causally efficacious, it follows that we should say, for example, that the loudness of the shot is relevant to whether the shot causes a death or not. Marras claims: “On Davidson’s account, the loudness of the shot, or even its having been fired from a dusty gun, would be as relevant to the death it caused as the physical properties of the shot that we would normally hold responsible for the victim’s death….”¹³⁹⁰ This, however, cannot be the case according to Marras because: “Clearly, not all properties of an event are equally relevant, or essential, to what it causes.”¹³⁹¹ I would answer perhaps they are, perhaps they aren’t; it partly depends on what one means by the term “relevant”. If we consider what the “ultimate” answer to the question “What caused X?” was, we would perhaps have to include everything preceding X in a spatio-temporal region of a certain size. Davidson is not eager to analyze

¹³⁸⁴ Davidson, 1982a, 186.
¹³⁸⁵ In Caorsi, 1999, 327.
¹³⁸⁶ Davidson, 1982a.
¹³⁸⁷ Davidson, 1993f.
¹³⁸⁸ Davidson, 1993g, 287.
¹³⁸⁹ As Davidson (1999x, 573) notes, an insufficient explanation does not mean that nothing has been explained. The fact that we do not know how a certain physical mechanism behind mental causation works does not mean that explanations referring to mental causes would not be explanatory.
¹³⁹⁰ Marras, 1997, 185.
¹³⁹¹ Ibid.
causation in terms of counterfactuals because counterfactual claims should always be evaluated in the context in which they are made. This is just to repeat the point that explanations in general must be evaluated in a specific context. I believe that this suggestion of Davidson’s is not completely consistent with his claim that “[…] interest aside, every property of every event is causally efficacious.”¹³⁹² What does causal efficacy outside all interests mean? Could it mean that every property of every event can be chosen to be a causally efficacious property depending on our interests? It is not easy to conclude how this “hyper-essentialism” about events should be understood. Commentators like Burge and Sosa have found it implausible; Burge thinks that the thesis is indefensible.¹³⁹³ Davidson himself notes that although his view has the consequence that, say, if a particular soldier in the First World War had not been shot then the First World War would have not been the First World War, the question of whether this would make any difference depends on the context where this counterfactual claim is made. Because counterfactuals depend on context, it is a mistake to base our metaphysics on them; Davidson thus rejects the counterfactual analysis of causation. This being said, he claims that laws are true generalizations which support counterfactuals precisely because the correctness of laws and explanations, like counterfactuals, depends on our interests.

Although it is true that one could argue that the dustiness of a gun could be relevant to what it causes, it does not follow that on Davidson’s account we should think that the dustiness of the gun is a relevant property to which we should refer in explaining why the shot caused a death. Remember: “What we call ‘the’ cause of some event is some feature chosen from the…causal factors which interests us…”¹³⁹⁴ and “[…] depending on the sort of explanation we are interested in, different properties of events are treated as causally efficacious.”¹³⁹⁵ It seems to me that those who argue that the dustiness of a gun is a relevant property that should be considered if we want to find out what was it about a shot that caused a death neglect the fact that causally efficacious properties are derived from explanations which satisfy our particular interests. It is an obscure claim that the properties which seem to be irrelevant for the occurrence of a certain effect would be as relevant as those which we actually take to be relevant. When we want to explain why a certain effect occurred we treat certain properties as being more relevant than others. There is no “metaphysical answer” to the question of whether they are more relevant. As far as mental causation is concerned,

¹³⁹² Davidson, 1987a, 113.
¹³⁹⁴ Davidson, 1993g, 287, my emphasis.
¹³⁹⁵ Davidson, 1987a, 113, my emphasis.
another way to argue against Marras could be the following. Certain properties matter to those causal relations which ground rational explanations. Mental predicates in terms of which rational explanations are given supervene, above all else, on behavioral predicates. They do not supervene on all physical predicates of an object. What is relevant for the explanation – and thus causing – of action are those features which supervene on the causal interactions between agents and events and objects in the world, and not the features which supervene on neurophysiological and physical facts. Davidson notes that the supervenience of mental on physical facts of various sorts does not suggest epistemological priorities. As far as I can see this cannot be entirely correct; on a Davidsonian account the behavioral level is privileged, and if we take von Wright’s argument seriously then this level should be privileged.

When Davidson talks about causation it is often causal explanation that he is actually referring to, although he also wants to keep questions about causation and causal explanation clearly separate. We can take a “generous attitude” towards the nature of cause and then all kinds of things can be causes. Stones can be causes, people can be causes and states can be causes, as Davidson claims. So it is not the case, as Crane for example claims, that “followers of Davidson will say that only events can be causes…” Crane, who vigorously wants to have causally efficacious properties, thinks that followers of Davidson cannot say that a skater’s weight or the ice’s fragility is literally a cause of the ice’s breaking. But what does “literally” mean? When the Davidsonian generous attitude towards causes is taken, it certainly can be said that the weight of the skater caused the ice to break; to insist that this was literally the cause does not make sense from the generous point of view. The generous attitude agrees with the way the concept of “cause” is used in ordinary speech and in this context, so it seems to me, the concept is very closely tied to that of explanation. Insofar as we want to talk about mental causation, or insofar as Davidson uses this expression, the talk should be understood as referring to mental explanations because the expression “mental cause” is clearly tied to our ordinary notion of cause. This being said let us draw a conclusion of Davidson’s views about mental causation by briefly considering the peculiarities of rational explanations and how they differ from physicalistic explanations.

In section 3.2, the rationalizing aspect of rational explanations was already discussed. I claimed, drawing strongly on the views of von Wright, that the explanatory force of rational

\[ ^{1396} \text{Davidson, 1983.} \]
\[ ^{1397} \text{Davidson, 1993g.} \]
\[ ^{1398} \text{Crane, 2008, 181.} \]
explanations is not based on the assumption that a causal relation between attitudes and action exists; the derivation goes the other way around. If we think of reasons as causes it is because we have taken a generous attitude towards the term “cause”. Although Davidson in “Actions, Reasons and Causes” claimed that rational explanations are, and must be, a form of causal explanation, it should now be clear that the expression “causal explanation” must be understood differently in each case. Davidson himself seemed to loosen the idea that reasons are causes, emphasizing instead that they are rational causes. What is obvious is that the explanatory force of rationalizations comes from a different source than the explanatory force of causal explanations. “Mental causality” is a fall-out of rational explanations, which in turn should be understood as a form of understanding explanations in roughly the sense that is familiar from the views of von Wright.

Although Davidson sometimes uses the term “mental causation”, we should conclude that this expression is confusing from a Davidsonian perspective. Events cause other events, but the fact that an event happens to be a mental event does not explain its causal powers. The fact that the event is mental may, however, explain why we can explain something by invoking the mental description of the event. This being the case, it can be said that mental events cause the effects which can be explained by referring to mental descriptions. More than Davidson von Wright wants to make a clear distinction between reasons (which belong to the realm of mental, or to the “world of spirit”) and causes (which belong to the realm of laws). He notes: “[…] the determinants of action, I would maintain, are of totally different kind from causes and effects among events in nature. They fall under a different concept of causation…”

Interestingly, Davidson observes that: “The fact that these relations [between the mental and the physical] are so deeply puzzling suggests some sort of misfit between the concept of causality and the concept of a thought.” This is an interesting view when compared against the main claim of “Actions, Reasons and Causes” which was, after all, that reasons, which belong to the realm of the mental, must be causes. But it seems to me that the “misfit” now suggests that reasons cannot be causes in the strict sense of the word. As Davidson admits: “[the ordinary notion of cause], which is closely tied to explanation, and is therefore sensitive to how cause and effect are described, differs radically from the notion of cause we employ when we are describing a relation in nature between events.” In Davidson’s view there are two fundamentally different notions of “cause”; it is easy to see the

---

1399 Von Wright, 1974a, 2.
1400 Davidson, 1993f, 298.
1401 Davidson, 1993f, 307. As I have noted it is very unclear what the “relation in nature” can be on a Davidsonian account.
similarity of this view to the view of von Wright. Although “Actions, Reasons and Causes” has been usually interpreted as defending a position which brings reasons and causes together, I believe that the irreducibility of the mental marks a clear conceptual distinction between them. A naturalist would applaud “Actions, Reasons and Causes” while thinking that “Mental Events” and its defense of anomalism and irreducibility of the mental spoils the picture.

It therefore seems that when we speak of mental causes on a Davidsonian account we are not, strictly speaking, implying that the mental phenomena to which we refer are causes. Therefore critics who accuse Davidson of epiphenomenalism have a point. A widely recognized worry is that Davidson’s views have epiphenomenalist and antirealist aspects. These worries are not completely misplaced – although, as I have argued, some critics seem to derive these worries from the wrong sources – for example, from the fact that Davidson does not explain how properties could be causally efficacious. In fact, Davidson does offer a proposal about what makes a property causally efficacious – namely, its occurrence in a law by reference to which explanations are given. This is obviously not a kind of answer that would satisfy the critics, because they want to know what it is about the properties that makes them causally efficacious. But it should be also obvious that a theory like Davidson’s simply cannot provide a detailed account of mental causation. Why? Because the main claims of AM are: 1) from an ontological point of view mental phenomena are identical to physical phenomena, and 2) explanations referring to mental phenomena are autonomous. If we could explain how mental phenomena cause physical phenomena, or vice versa, we would run the risk that mental explanations could be replaced by physical explanations, thus making the former useless. This possibility is real because mental phenomena are physical phenomena, and therefore there exist physical relations between the phenomena which fall under strict laws. So, given Davidson’s views about the nature of the mental and its irreducibility, it is not surprising that he does not and cannot offer a theory of mental causation. This being said, it could be claimed that Davidson does offer a half-hearted theory of mental causation. It is a brute fact about events, mental or physical, that they have causal powers. But this answer does not satisfy a critic who wants to know whether the mental aspect of an event has causal efficacy.

Stoutland (1999) has argued that the difference between von Wright and Davidson is largely terminological. Given Davidson’s views as described here, Stoutland’s claim seems to correct.

I thank Louise Antony for sharing this personal impression of hers about Davidson’s position.
Since Davidson does not explain how mental causation works, his position could be accused of involving epiphenomenalism. He admits: “[…] we can’t say in clear detail how the mental attitudes cause the action”.\textsuperscript{1404} This kind of view does not make one’s position epiphenomenalist. But what about the claim which was mentioned already: “There is a harmless sense in which one can say the contents of the reasons on which we act is irrelevant to the causation of the action, because we can, in theory, explain why the action occurred using only physical laws and descriptions.”\textsuperscript{1405} The way I interpret this view is that something causes an action and the content of the reason which explains the action is irrelevant to the causation. This sounds like a form of epiphenomenalism. Further interesting remarks make it problematic to say what Davidson actually thinks about the causal efficacy of reasons. He notes: “[…] we [cannot] infer from the fact that what we typically think of as the reasons for an action are not events that there aren’t events that are essential to the causing of the action. These events may or may not be mental in any ordinary sense.”\textsuperscript{1406} So, the reasons to which we refer as the reasons for action are not necessarily the causes of the action; in fact the event which causes the action may not be mental in any ordinary sense. What causes my action can be something other than the mental event by reference to which I explain the action. This being the case, it should be obvious that Davidson’s references to “mental causality” should really be taken metaphorically. It is not the case, at least not obviously, that the reasons to which we refer as reasons for actions pick out causally efficacious physical events. This is not surprising given Davidson’s “interpretationism” and its consequence that explanatory attitudes need not correspond to anything running through the agent’s head before an action takes place.

I think that the most severe problem is that “Davidson’s reasons” may be both non-existent and non-efficacious. One could argue that there is therefore a very easy route to epiphenomenalism in Davidson’s theory – namely, a kind of antirealism about mental events. If the existence of mental states is doubtful then so is certainly their causal efficacy. In fact, as I have shown, Davidson concludes that talking of beliefs, desires, and intentions as causes and effects is a metaphorical. This is because, \textit{strictly speaking}, only \textit{events} can be causes and mental “entities” like beliefs are not events. Davidson claims: “[…] when we mention beliefs and desires to explain an action, we are describing key aspects of the circumstances under

\textsuperscript{1404} Davidson, 1993f, 307.  
\textsuperscript{1405} In Caorsi, 1999, 338.  
\textsuperscript{1406} Davidson, 1999, 654, my emphasis.
which the agent acted.”¹⁴⁰⁷ Or, “What we mean, in my view, when we speak of beliefs, desires, and intentions as causes of actions is that they are significant causal conditions of actions.”¹⁴⁰⁸ These remarks suggest that Davidson does not see reasons as causally efficacious entities and this suggests that the explanatory force of rational explanations should be seen following from the fact that they enable us to see and understand a situation they explain in a certain way.¹⁴⁰⁹

An interesting consequence of this is the following. Davidson’s most important claim in the philosophy of mind is that mental explanations are autonomous and indispensable. Against this view, the claim that we could, in theory, explain why an action occurred using only our physical vocabulary is obscure. However, since Davidson also argues that “no one” would accept the physical explanation as an explanation of action viewed as an action, the irreducibility of reason-based explanations stands. Therefore it is difficult to see the point in claiming that “in theory” an action could be explained with only physical concepts. Davidson seems to think of actions in two different ways. On the one hand, they are physical phenomena and therefore there is, in theory, a physical description which applies to them. On the other hand, they are mental phenomena and identifiable as actions only in our mental vocabulary. But if the irreducibility of the mental operates also at the token level, why would we be willing to hold on to the idea that actions are in any clear sense physical phenomena?

4.2 The new problem of mental causation, epiphenomenalism and precursory reasons

I have suggested that if the irreducibility of the mental is accepted, then a reductive explanation of mental causation is impossible.¹⁴¹⁰ Sometimes Davidson expresses a positive view about the possibility of certain kinds of reduction – as for example, when in one of his earliest articles he claims: “[…] there is no reason why we cannot say this particular event (and here we describe it with a physical description) is a causal factor in the production of this mental event.”¹⁴¹¹ Davidson seems to be ambivalent about the question of what science can and cannot explain about mental phenomena. The same is true of von Wright, who claims that the problem of consciousness is a philosophic muddle which cannot be clarified through science, but who also refers to scientific results as a reason to believe in epiphenomenalism. It

¹⁴⁰⁷ Davidson, 1993g, 287-288.
¹⁴⁰⁸ Davidson, 1999q, 499.
¹⁴⁰⁹ See section, 3.2.
¹⁴¹⁰ Hutto (1999) has reached a more general conclusion: If psychology must be understood by interpretationalist lights, then it simply cannot be reduced or naturalized.
¹⁴¹¹ Davidson, 1964, 48.
seems to me that even Davidson’s “positive remarks” about the prospects of science go too far if the claims about token-irreducibility and of the nature of mental states are taken seriously. Indeed, as Davidson notes in his skeptical mood: “[…] human thoughts correspond to so little of the material world that no self-contained science can be based exclusively on them.” \(^{1412}\)

In chapter three I gave reasons to believe in the irreducibility of the mental and showed the extent to which Davidson and von Wright are committed to the irreducibility of the mental. When discussing these somewhat technical questions, the most important thing is to recognize the “spirit” in terms of which von Wright and Davidson think about the place of the mental vis-a-vis physical reality. In the last three sections I have described Davidson’s relation to the problem of mental causation and considered how he could answer the problem. I conclude that from the perspectives of Davidson’s and von Wright’s positions, a reductive solution is impossible. Let us nevertheless note that Davidson makes an interesting although obscure remark about the possibility of a solution:

\[
[\ldots] \text{the problem of how beliefs and desires cause an action when they give the reason it was performed; the problem of how external events cause sensations and beliefs in perception and memory; the problem of how one belief causes another when we reason – could it be the case that there exist “solutions” to these problems, even if for some reason we cannot arrive at them?}^{1413}
\]

The claim that there could be solutions although we could not arrive at them is puzzling. We could only speculate about what kind of “solutions” Davidson might have had in mind.

I believe we have to once more ask what the problem of mental causation actually is. Kim states that the problem is not so much one of whether mental causation is possible, but how it is possible given certain assumptions about the nature of reality. He rejects positions which claim that mental causation is an obvious fact and that there is no special problem related to it. Kim admits that the existence of mental causation is a fact, but this does not remove the worry about how it is possible. I want to argue that the how-question should be considered in a broader setting than in the context of O-physicalism. I have suggested that the new problem of mental causation is the question of how to tell that a mental explanation is true. The problem of mental causation can be approached either from a metaphysical perspective where the essential question is how an overdetermination of action by mental and physical causes is possible, or from a methodological one where we ask what the nature of mental explanations is and how they differ from scientific explanations. It would be an

\(^{1412}\) Davidson, 1993f, 312.
\(^{1413}\) Davidson, 1993f, 313.
exhaustive task to address the contemporary discussion, which accepts as its starting point the metaphysical perspective. My intention is to consider how the problem of mental causation could be thought about outside the jungle of empirical – physicalistic – metaphysics. There are three reasons for this. The first is the attempt to clarify how the problem could be addressed from Davidson’s and von Wright’s points of view. The second is the belief that there are good reasons to think that the irreducibility of the mental is true. The third is the conviction that a reductive solution to the problem of mental causation should not be sought because it jeopardizes the distinctiveness of human agency.

The modern way to state the problem of mental causation seems to start from two assumptions. The first is that epiphenomenalism is false. The second is that mental properties are real. Given these two assumptions, the problem is to explain how mental properties are causally efficacious, that is, how mental events or states cause something “qua mental”. I do not wholly understand how a contemporary physicalist can formulate the problem in this way. If it is agreed that reality is thoroughly physical, then the expression *qua mental* is misplaced from the start. If causation is a physical phenomenon and if there is such thing as mental causation, then the question is what kind of physical states mental states are, and what allows them to be causally efficacious? I think that in order to answer this question we would have to show how the mental reduces to the physical by explaining how the “entity” that we now describe in our mental vocabulary could be described in our physical vocabulary. If we could describe this and if we could describe how causation worked between physical entities, we would have answered the question of “How mental causation is possible” – without invoking *qua mental* expressions. This is how the situation should be from a properly physicalistic perspective, because in a physical reality there is not, or at least should not be, room for things to cause anything “as mental”. From the perspective of O-physicalism, both proposals 1) and 2) from section 4.1 will do as solutions to the problem of mental causation; they merely mark the difference between metaphysical and methodological approach. If we are metaphysically oriented physicalists, we should proceed in the spirit of suggestion two, which is the line taken for example by Kim. If we are more epistemologically oriented, I believe we should explore the possibilities of suggestion number one. Keeping these possibilities in mind I want to suggest a different kind of solution, which is based on the view that the possibility of epiphenomenalism should be taken seriously. It seems to me that our inability to provide a reductive solution to the problem of mental causation opens up this possibility. Whereas a reductive solution could kill the patient in the process of seeking a cure, it could also “vindicate” mental causation.
In considering the possibility of epiphenomenalism let us consider how the problem of mental causation could arise. We have observed action A and we think that a mental state M caused A. The situation is like this:

\[ M \rightarrow A \quad (M \rightarrow M) \]

The arrow implies a causal relation. Given what we think we know about human physiology we also believe that there is a physical cause for the occurrence of A. The situation is like this:

\[ P \rightarrow A \quad (P \rightarrow M / P \rightarrow P) \]

The arrow stands for a causal relation between a physical state P and action A. On the one hand there seems to be a causal relation between mental phenomena (M – M); on the other hand there seems to be a causal relation between physical and mental phenomena (P – M).\(^\text{1414}\)

The problem of mental causation arises when we ask what the relation between these two relations is. Are there two relations or just one? The problem is easily seen if we assume that everything that happens must have a sufficient physical cause. Then it seems to be the case that the relation P – M is the only required causal relation. This raises the question of what the role of M – M relation is? The problem is to explain what the relevance of M – A is, if the causal relation obtains between P and A.

When discussing how supervenience is related to the problem of mental causation, Kim claims that there are two possible answers to why M*\(^\text{1415}\) was instantiated on a certain occasion. Either: a) M caused M*, or b) P*, the physical supervenience base of M* is instantiated. Kim claims that there is a real tension between these answers and it is this: in order for M to cause M*, M would have to cause P*. This would be a case of mental-to-physical causation, which is not easily tolerated if the truth of supervenience is assumed. In this context supervenience is defined by Kim in the following way:

Mental properties supervene on physical properties in the sense that if something instantiates any mental property M at t, there is a physical base property P such that the thing has P at t, and necessarily anything with P at a time has M at that time.\(^\text{1416}\)

---

\(^{1414}\) If we understand the outcome as an action. Otherwise the relation is between physical phenomena (P – P).

\(^{1415}\) In Kim’s formulation M* is some mental property. We can here think of it as referring to A in our example.

\(^{1416}\) Kim, 1998, 39.
Given this assumption, it follows that M has a physical supervenience base P. From this it follows that in order for M to cause M*, M has to cause P*. It should be noted that this last assumption is based entirely on Kim’s metaphysical understanding of supervenience. According to this view, higher level properties depend on the properties at a lower level – and this being the case a base property must be instantiated if a supervenient property is to be instantiated. As Kim claims when giving an example of how aesthetic properties supervene:

To make your painting more beautiful… you must do physical work on the painting and thereby alter the physical supervenience base of the aesthetic properties you want to improve. There is no direct way of making your painting more beautiful or less beautiful; you must change it physically if you want to change it aesthetically – there is no other way.1417

But is this really so? Does a “physical supervenience base of aesthetic properties” make sense? It seems that, according to Kim, an aesthetic property is something which can be clearly and in some sense observer-independently identified. But do we know what kind of properties aesthetic properties are? Could it not happen that a painting looks more beautiful to me today than it did yesterday without there being any relevant physical change in the painting? Would it be false to say that, this being the case, the painting isn’t more beautiful? Kim argues that the physical properties of a painting must be changed in order to change its aesthetic properties; “there is no other way”. But why could we not change the person who evaluates the painting? This, of course, would not make the painting more beautiful. But are there such things as intrinsically beautiful paintings? Is it correct to claim that aesthetic properties are independent of observers? If not, then it is not just the properties of the painting on which its beauty supervenes; this can be likened to Davidson’s later understanding of supervenience. The supervenience base is wider than the intrinsic properties of an object – recall the example of a sunburn.

What are the facts or physical properties on which the beauty of a painting supervenes? This is a question which is not easily answered. If we consider why this painting is beautiful now although it wasn’t beautiful one month ago, it is not clear what the total supervenience base is that should be considered. It is perhaps something in the painting, although it cannot be anything which wasn’t there before. Perhaps one’s knowledge of the painting has increased and thereby it somehow seems, or as I would say is, more beautiful. Or perhaps the observer is just in a happy state of mind and therefore “everything seems beautiful”. These are not naïve answers to the question “Why is this painting beautiful?”.

1417 Kim, 1998, 43.
because we cannot make a clear distinction between something *seeming* beautiful and something *being* beautiful. Suppose I say that the painting is more beautiful today than yesterday. If supervenience is understood as a non-reductive relation, then it is not possible to say what the properties of the painting are, on which its beauty supervenes. But if this is true, how could I insist that the painting *must* have changed? Can we exclude the possibility that the beauty of the painting depends entirely on me? Yesterday I could not see the beauty of the painting; today I see that the use of this color, the arrangement of these lines is the reason why the painting is beautiful; now I see its beauty. But no physical work has been done on the painting.

If this is the case also with mental properties, that they are in a sense interpreter-dependent, then it is not clear that we could make sense of mental properties which could be independently identified and thereby clearly correlated with their physical base. If the situation is seen from the perspective of “Davidsonian supervenience”, we do not have to assume that the supervening properties of the painting would depend just on its *intrinsic properties*. If the supervenience relation is to be understood as holding between *predicates*, then we can say that there has to be some physical change if the painting is to change from grotesque to beautiful – but the change does not have to occur in the painting. If, instead of understanding supervenience as a metaphysical thesis, we understand it as a thesis about our use of words or about the logical or conceptual relations between the application of certain concepts, then we do not have to insist that a change in the supervenient property (predicate) is linked to a narrow physical supervenient base through laws.

Now we can ask what the causal status of M and P with respect to P* is, and thereby to M*. Kim argues that since P is sufficient for M and M is sufficient for P*, P is sufficient for P* – and we thus have a reason to claim that P *precludes* M as being a cause of P*. Kim claims that “the most natural way” to see the situation is to say that: P caused P*, and M supervenes on P and M* supervenes on P*.\(^{1418}\) Then the situation looks like this:

\[
\begin{array}{c}
\text{M} \\
\uparrow \\
\text{P} \\
\rightarrow \\
\text{M*} \\
\uparrow \\
\text{P*}
\end{array}
\]

\(^{1418}\) Kim, 1998, 45.
The arrows from P to M and from P* to M* are best understood not as causal relations but perhaps as relations of “constitution” or “realization”. The problem for the reality of mental causation now is that there seems to be only one causal relation, and that exists between physical states. What happened to M – M, P – M, or M – P causation? The source of the problem is obviously the assumption that there is such thing as causation where one member is a mental phenomenon. I think that Kim is correct in his view that mental causation does not fit the picture described by non-reductive physicalism very well. In my opinion, the reason for this is that the assumptions of non-reductive physicalists about the nature of causation are thoroughly physicalistic. What follows if we challenge these assumptions and start from the claim that it is not obvious that there is any kind of causation in which one member is a mental phenomenon?

The consequence of this would be something which could be called the causal impotence of thoughts. This suggestion may seem implausible. Depending on how strong we think the claim to be, it turns out that: a) mental phenomena never cause anything (which is the standard understanding of epiphenomenalism), b) mental phenomena cause something only rarely, or that c) we have no way to tell when mental phenomena cause something, and this being the case we can never be sure, on a given occasion, whether mental phenomena were causally efficacious. The last option opens also up the possibility that in each specific case mental phenomena are causally inefficacious. My “epiphenomenalist solution” does not however claim that mental phenomena cannot ever cause anything. This suggestion would be too harsh since it would so blatantly conflict with our common-sense understanding. My claim is rather that in many cases thoughts do not cause the things we take them to cause – in these cases thoughts are causally impotent – and if we choose to think that thoughts did cause something, our only evidence for this is a mental explanation which we decide to trust.

### 4.2.1 Epiphenomenalism

Although epiphenomenalism has been seen as an alternative in the history of philosophy it is not currently a widely accepted view. Putnam, for example, has labeled it as a C-R-A-Z-Y doctrine. The status of epiphenomenalism and the possible reasons for it have nevertheless

---

1419 It is of course possible, in some sense of the word, that our common-sense understanding is completely false, but if this is really the case then this really seems to be the end of the world as Fodor (1987) has noted.

1420 Putnam, 1999.
been under discussion also recently. Suggestions especially referring to neuroscientific and psychological evidence tentatively point towards the conclusion that perhaps a form of epiphenomenalism could be true after all. These suggestions seem to be better accepted among neuroscientists than among philosophers.

Global epiphenomenalism, which claims that mental phenomena are always causally impotent never cause anything, is an inconvenient conclusion. At the same time it has to be admitted that this, the total causal inefficacy of mental phenomena, is possible and it may actually be our situation. There is no evidence which could settle the question of whether global epiphenomenalism is true or not. It may be the case that mental phenomena never cause anything and that a kind of pre-established harmony is true, but we will never know whether this is so. Let us therefore start with the assumption that global epiphenomenalism is false, because mental phenomena seem to have causal efficacy. But can the causal efficacy of mental phenomena be proven? Was a specific mental phenomenon really a causal factor at time t? Did it cause the action which is explained by reference to it? A reductive explanation of mental causation in physical terms would be a step towards a vindication of mental causation. This way is blocked for those, like Davidson and von Wright, who defend the irreducibility of the mental. My claim is that the possibility of conditional epiphenomenalism threatens those positions which claim that an actual reduction of mental phenomena to physical phenomena is not possible. According to the doctrine of conditional epiphenomenalism, mental causes, which we take to be causally efficacious, are causally inefficacious from time to time. If conditional epiphenomenalism is true, then it is always possible that on a specific occasion a mental phenomenon in terms of which we explain our own actions or the actions of others is without causal power. Whereas global epiphenomenalism may be labeled a crazy doctrine and its interest for our lives is minimal, the possibility of conditional epiphenomenalism is something which has importance for various issues. If conditional epiphenomenalism is true then the questions, for example, about the status of mental explanations, responsibility, or self-knowledge have to be reconsidered. If a mental cause can be causally inefficacious on any specific occasion, how to decide whether an agent is to be blamed or praised of what (s)he did? If a mental cause can be epiphenomenal, then how to determine whether I acted when I thought I acted? If there are

---

1422 This being the case we cannot even consider which position - global epiphenomenalism or non-epiphenomenalism - is more probable.
1423 If global epiphenomenalism is true then the causal efficacy of thoughts cannot be proven.
two candidates for a mental cause in terms of which I explain my action, how to decide which one is causally efficacious?

Because of these important consequences, the possibility of conditional epiphenomenalism should be taken very seriously. But are there reasons to believe that this form of epiphenomenalism is actually true? Indeed there are. Non-reductive physicalism is especially vulnerable to the threat of conditional epiphenomenalism, because in its context the causal efficacy of mental phenomena cannot be proven by showing that they are identical to the states of brain (which we assume to be causally efficacious). But to say that non-reductive physicalism is vulnerable to the charge of epiphenomenalism or, as many contemporary philosophers claim, that it leads to epiphenomenalism, is misleading because conditional epiphenomenalism is a general challenge. Conditional epiphenomenalism is just not a philosopher’s fantasy but a position which may describe our actual situation; there is sound evidence that we are often mistaken in self-reports about the reasons for our actions. What, then, are the reasons for believing that conditional epiphenomenalism is true? I will consider von Wright’s reasons for defending epiphenomenalism. In doing this, I will briefly refer to scientific evidence which supports the view that conditional epiphenomenalism could be true. The focus will be on evidence that is the result of recent studies of the brain. This kind of evidence has led philosophers to consider the possibility of epiphenomenalism. In addition to the results of neuroscience, I will also consider evidence from studies in psychology.

4.2.1.1 Von Wright’s epiphenomenalism

The problem of mental causation troubled von Wright. He noted that it had always been a mystery to him how a causal mechanism between a physical event and, for example, a sensation could work, and concluded: “[…] no discoveries of a scientific nature, however exciting, would make me understand better, I am afraid, the ‘transformation’ which occurs when a sensation originates.”1424 Perhaps it was this mysterious connection which led von Wright to claim that in the picture of natural connections, the mental has no role. In the part of reality investigated by science, the mental has no place. This is an intriguing suggestion. Von Wright seems to claim that the problem of consciousness and the problem of mental causation will never be solved scientifically. I believe he would thus disagree with, for example Kim, who claims that there will be no special mystery of mental causation if mental

1424 Von Wright, 1997a, 152.
phenomena are neural processes in the brain. Reduction provides a direct and simple account of mental causation. Von Wright thinks that this would not dispel the mystery around the problem.

Von Wright’s reasons for arguing for epiphenomenalism give us one interesting way to think about the problem of mental causation. Unlike most contemporary philosophers von Wright admits his affiliation to epiphenomenalism. This is surprising; epiphenomenalism is a strange position and not many philosophers are willing to defend it. Von Wright was however tempted by epiphenomenalism and wrote that: “In the view that I am... advocating, there is no causal interaction between the mental and the material.” This is a straightforward claim. Von Wright is willing to make an even more daring claim when he notes that the denial of mental–physical interaction is essential to his position. Although, as we will see, this claim can be given a weak reading, von Wright nevertheless defends dramatic intuitions when he argues that consciousness stays outside the spatio-temporal web of causally related events, or that a sensation is neither caused nor causally efficacious. The latter claim von Wright affiliates with epiphenomenalism as it is traditionally understood. This suggests that von Wright accepts the central principle of epiphenomenalism. It could be argued that he goes even deeper to the maze of epiphenomenalism than traditional epiphenomenalists like Huxley who denied the causal efficacy of mental phenomena but were not willing to claim that mental phenomena are not caused by physical phenomena. Von Wright namely suggests that sensations cannot cause physical phenomena and cannot be caused by them. Kim, who has claimed that Davidson’s position leads to the same consequences, has noted that “Cartesian souls” would be very lonely “entities” because they would be causally cut off from everything. If consciousness stays outside the spatio-temporal web, then von Wright’s Cartesian consciousness is in danger of becoming the kind of lonely soul described by Kim.

The denial of mental-physical interaction means that: “There is no... causality in the sense of either “material cause – mental effect” or “mental cause – material effect.” This bears a resemblance to the view that we have already considered, namely:

---

1425 The material of this section is taken partly from Kuusela, 2009.
1427 In the only review of In the Shadow of Descartes that I am aware of, Kivinen (1998) describes the claim as “extremely fudgy, to put it mildly”.
1428 Von Wright, 1997a.
1429 Ibid.
1430 Ibid.
A physical event P causes (C) another physical event P*. According to Kim, there is a relation of constitution or realization between P and M and P* and M*. In this picture, there is no room for mental causality. Von Wright’s claims that a relation exists between P and P*, but there is no relation of mental causation. There is, however:

[...] a connection or, if one wishes, parallelism between the two [mental and the physical]. This "parallelism" is there by virtue of the criterional ("semantic") relation between, on the one hand, behavioural and mental (intentional) phenomena and, on the other hand, mental phenomena (sensations) and things and events in the physical world.¹⁴³³

In the previous diagram, I have marked, along von Wrightian lines, with “S” (for semantic) the relation which Kim takes to be a metaphysical relation of constitution or realization. I think that von Wright’s suggestion fits well with how Davidson might think of the case if we understand Davidson’s version of supervenience as being a semantical relation.

Is this view plausible? It amounts to saying that when I have a reason R and I think that I act because of it, there is nevertheless no causal relation between my reason and my action. If epiphenomenalism is true, on what is our conviction that cases of “mental cause – material effect” are commonplace based? I think that von Wright’s answer is half-hearted. It is based on the view that in humans the development of cognitive capabilities goes hand-in-hand with the development of the brain. The congruence between the mental and the bodily aspects of action is a harmony established in the course of individual’s life. Despite this harmony, which is best described in terms of the semantic relation, there is no causality involved between the mental and the physical. The conclusion of von Wright is:

We have two sets of events in the world of space and time, overt behaviour and neural processes, and between them causal relationships. We also have, like a “shadow” accompanying the first set of events, a sequence of mental states and processes, semantically connected with it.¹⁴³⁴

According to von Wright this suggestion is also a solution to the problem of mental causation. We should note that the solution works only because it denies the existence of mental causation. This is how von Wright’s “emergentism” or epiphenomenalism avoids the problem of mental causation. Von Wright’s version of supervenience was discussed in section 2.2.1. His use of this concept illustrates how the neural level has causal priority. Von Wright writes: “A causal relationship between a mental and a physical phenomenon is supervenient in time on a causal relationship between two physical events…” A formulation which illustrates the causal role of physical events with respect to mental ones is the following:

Whenever a physical event P can be correctly called cause (effect) of a mental event M, there exists another physical event F such that the duration of M is included in the duration of F, and such that F is the effect (cause) of P.  

I claimed in section 2.2.1 that von Wright’s views about supervenience are not easy to understand. It is not clear what a claim that a “volitional act is temporally included in cerebral processes” means, or how we should understand the suggestion that “a reason is a mental episode which is a temporal segment of a neural cause”. In order to honor the honesty of von Wright, it has to be noted that he considered these views very difficult even for himself. I believe that using supervenience as a semantic relation could provide one clarifying way to describe von Wright’s view, although he himself did not use this expression.

Von Wright accepts a form of epiphenomenalism, but what is his reason to claim that the mental and the physical do not interact causally? The mere assertion that the relation between the mental and the behavioral is semantic is not enough. In von Wright’s case there are two reasons to argue for epiphenomenalism. First, by denying mental–physical interaction “conceptual convenience” is gained, because a conflict with the views about the workings of causality in nature is avoided. A denial of interaction safeguards the causal closedness of the physical world order. What is denied is the possibility of Cartesian interactionism, while at the same time a special form of interaction is affirmed. This is done by introducing a relation of supervenience which replaces the talk of mind–body interaction with talk of causal connections between physical events only. In this sense, von Wright accepts interactionism without violating the principles of physical science. Interestingly, von Wright describes this

\[1435\] Von Wright, UPd, 6.
\[1436\] Ibid.
\[1437\] Ibid.
\[1438\] A note expressing von Wright’s worries was attached to the unpublished manuscript.
position, which accepts interaction, as a position revindicating some Cartesian intuitions.\textsuperscript{1439} Acceptance of interactionism of this kind is also an acceptance of a form of epiphenomenalism given von Wright’s view that causal relations occur between physical events and that the mental is merely a byproduct of underlying neural phenomena. The situation can be illustrated with this diagram.

\begin{center}
\begin{tikzpicture}
\node (s) at (0,0) {S};
\node (n) at (1,0) {N};
\node (m) at (0.5,1) {M};
\node (b) at (1,1) {B};
\draw[->] (s) -- (n);
\draw[->] (n) -- (b);
\draw[->] (m) -- (n);
\draw[->] (m) -- (b);
\end{tikzpicture}
\end{center}

\textit{time}

S here stands for sound, N for a neural event, M for the sensation of sound, that is, hearing, and B for behavior. Two problematic relations of the diagram are the assumed to be the physical–mental relation between N and M and the assumed mental–physical relation between M and B.

The diagram describes a situation where a sound S causes a neural event N. A neural event in turn causes a bodily reaction B. This is a causal explanation involving only physical phenomena. \textit{Something physical}, a state or succession of states, takes place in the brain as a result of the sound. N can be understood to include everything physical in the brain which mediates between the sound and consequent behavior. N would include various neural states and processes. If everything which happens has a cause, it must be the case that if M (hearing) occurs, then S is also \textit{its} cause. A physical phenomenon S causes us to \textit{hear} a sound. To avoid a dualistic interpretation, M must be understood as a peculiar kind of neural state. In addition to the causal explanation we have also a reason-based explanation, which claims that M causes B.

N, understood as a succession of neural states, causes behavior B. What about M? M and N are a “joint effect”, a mental–physical effect, of S, but the relation between M and N is, according to von Wright, best understood as a relation of temporal simultaneity so that M is \textit{part} of N. M does not have to exist during the totality of successive states which form N, it may occur anytime between N, understood as a combination of N\textsubscript{1}…N\textsubscript{n}, and B. As we have

\textsuperscript{1439} Von Wright, UPd.
seen, it is not entirely clear how von Wright sees the relation between N and M. M is an emergent byproduct of N and the notion of supervenience is meant to clarify their relation. As noted in section 2.2.1, von Wright’s views about supervenience remain vague by being formulated in terms of temporal instead of spatial coincidence – and since he does not discuss the related concepts of “realization” or “constitution” the concept, just like Davidson’s “anomalous supervenience”, does not explain much.\textsuperscript{1440} The best description of the relation of M and N is von Wright’s claim that it is a part–whole relation, which cannot easily be understood as a relation of causality or a relation of identity. In the previous diagram, the dotted line between M and N illustrates this.

Von Wright argues that whereas N causes B, the part of it which is M is causally inefficacious. In order to show why the situation gives the appearance of mental cause – physical effect, we re-draw the diagram in the following way.

\begin{center}
\begin{tikzpicture}
\node (S) at (0,0) {S};
\node (N) at (1,0) {N};
\node (B) at (2,0) {B};
\node (M) at (0,1) {M};
\node (R) at (1,1) {R};
\node (A) at (2,1) {A};
\draw[->] (S) -- (N);
\draw[->] (N) -- (B);
\draw[->] (M) -- (R);
\draw[->] (R) -- (A);
\draw[dotted] (M) -- (N);
\end{tikzpicture}
\end{center}

Because of the occurrence of a causally inefficacious sensation which is simultaneous with a causally effective neural state, we interpret the behavior B as action A. This we do because M functions as a reason for us to act. But this does not make the sensation the cause of A / B. The causal relation, which we thought to exist between M and B, has been changed to an explanatory relation that explains M as a reason for A or A as a consequence of M. As we have seen, one way in which reason is defined by von Wright is as “[…] a mental episode which is a temporal segment of… physical cause of a bodily movement, which the reason is said to motivate.”\textsuperscript{1441} The relation between M and A can be understood in two ways. From the perspective of M, it is an explanatory relation which, so to speak, refers to the future and anticipates the occurrence of A. Seen from the perspective of A, it is a semantic relation which means that M occurred. We can understand action in terms of reasons by looking back after the action has taken place, or we can look forward from the perspective of reasons and

\begin{footnotesize}
\textsuperscript{1440} What further confuses things is the fact that when discussing supervenience von Wright uses expressions like “included” which certainly have spatial connotations.
\textsuperscript{1441} Von Wright, UPd, 5.
\end{footnotesize}
consider what is going to happen. Needless to say, if M is causally inefficacious it becomes puzzling how we could predict anything in terms of it. Von Wright does not discuss this important problem. Kim, who now doubts his own model of epiphenomenal or supervenient causation, claims that inserting the kind of arrow like R: “[…] could… be philosophically pernicious if it should mislead us into thinking that we have thereby conferred on M, the mental, event some real causal role.”\textsuperscript{1442} As I have shown, this quasi-causalism of M is also criticized by von Wright. But if M does not have causal powers, the pressing question is why a reference to it would have any explanatory power either. A partial answer to this problem will be given in the next section.

Are there reasons to think that this speculative model could reflect what actually happens at the causal level in the cases in which we understand behavior as action because we interpret the former as being caused by underlying mental states? It is certainly imaginable that something like this could be happening, but there is also empirical evidence to support this speculation. The second reason which points towards an epiphenomenalist conclusion in the views of von Wright is his interpretation of the empirical evidence given by neuroscience. He mentions especially the much debated work of Benjamin Libet.\textsuperscript{1443} One conclusion of Libet’s research is that a conscious decision to act, which in this context we may define as a reason, is preceded by an unconscious buildup of electrical charge, which we may define as the cause, within the brain. The finding of this \textit{readiness potential} is what has partly motivated the technologically oriented research that aims to develop brain–computer interfaces and related machines. Readiness potential, at least according to some neuroscientists, is the real cause of action. Already in 1983 Libet and his colleagues concluded that: “These considerations would appear to introduce certain constraints on the potential of the individual for exerting conscious initiation and control over his voluntary acts.”\textsuperscript{1444} Libet’s conclusions have been recently verified and further studied by neuropsychologist Angela Sirigu.\textsuperscript{1445} Desmurget, Sirigu and others have very recently claimed that:

\begin{quote}
[...] motor intention and awareness are emerging consequences of increased parietal activity before movement execution. The subjective (and potentially illusory) feeling that we are executing a movement does not arise from movement itself, but is generated by prior conscious intention and its predicted consequences.\textsuperscript{1446}
\end{quote}

\textsuperscript{1442} Kim, 2003, 148.
\textsuperscript{1444} Libet et al., 1983, 641.
\textsuperscript{1445} Sirigu et al., 2004, Kilner et al., 2004, Desmurget et. al, 2009.
\textsuperscript{1446} Desmurget et al., 2009.
In their study Desmurget et al. stimulated a part of the brain of a patient thereby creating a strong intention and desire for hand, arm, or foot movement. When the intensity of the stimulation was increased the patients believed they had really performed these movements. On the other hand, stimulation of the premotor region triggered contralateral movements but the participants firmly denied that they had moved. From these results the researchers noted: “[…] the fact that patients experienced a conscious desire to move indicates that stimulation did not merely evoke a mental image of a movement but also the intention to produce a movement, an internal state that resembles what Searle called ‘intention in action’”.

For a brain-oriented philosophy of mind driven by naturalism these findings are not surprising. It does create problems for a view, such as Davidson’s, according to which intentions are not “little entities lodged inside the brain”. If intentions can be artificially produced by stimulating the brain, then a claim that: “We do not expect to be able to change people’s beliefs and desires by tinkering with their brains” faces serious trouble. The recent findings of Desmurget and others certainly create interesting questions for the future and widen the gap between philosophically and scientifically motivated theories of mind. I think it is quite puzzling, definitely thought-provoking and possibly alarming that intentions are already produced artificially. Many who are enthusiastic about the prospects of neuroscience always insist that scientific discoveries cannot jeopardize free will and human freedom; they are not a threat. Philosophers who foolishly believe in compatibilism reach the same conclusion. I think it is quite obvious that if desires can be created simply by stimulating the brain and if we know that such and such stimulation causes such and such desire, then this will have some consequences for our autonomy and especially for our idea of autonomy.

In section 3.2 I briefly mentioned that the kind of scientific results described here perhaps challenge a philosophical non-reductive theory of action and its autonomous rational explanations. Von Wright, however, seems to use this kind of evidence, above all else, to defend his epiphenomenalist position. He argues, referring to neuroscientific findings, that if it can be shown that B occurs before M: “[…] this may be regarded as a ‘proof’ that the sensation (consciousness, something mental) is causally inefficacious i.e. it has no causal role to play at all.” He gives an example. Pain behavior may occur before pain is felt. This being the case, the sensation of pain has no causal function to perform which could be independently established. The time-lag between behavior and sensation must of course be

1447 Desmurget et al., 2009, 813.
1448 Rorty, 1999, 580.
1449 Petit, 1999 raises this challenge for von Wright’s position.
very brief, otherwise the agent in question would be puzzled and not understand the “connection” between the two. In von Wright’s opinion the case where B occurs before M: “[…] shows that sensation… plays no causal role in the production of the movement… this reversal of the temporal order *corroborates my thesis* that the ‘mind’ has no (independent) causal role whatsoever in producing changes in the material world.” In my opinion the claim that mind has no independent causal role in producing changes is startling; the claim of an eliminativist would not be more daring.

It is interesting that in addition to the “conceptual convenience” von Wright refers also to empirical evidence. The reference is perhaps best explained by von Wright’s respect for a scientific world view. In 1966 von Wright had a brief debate with another Finnish philosopher, Raimo Tuomela. Tuomela claimed that von Wright’s position has the consequence that a mental occurrence cannot be a cause of behavior or the cause of the goal, which is a causal consequence of the behavior. Von Wright replied that in his writings he had not made this challenging claim, but perhaps Tuomela thought he saw this claim “between the lines.” Von Wright then goes on to admit that he is indeed inclined to think that the claim is correct although he does not know how to prove it; the reason that von Wright is inclined to accept the view is that denying it would go against a healthy scientific world view. It can thus be seen that already in 1966 that “conceptual convenience” was von Wright’s reason for defending a form of epiphenomenalism. I suggest that, thirty years later, the reference to neuroscientific evidence is von Wright’s “proof” or “corroboration” for why the claim should be accepted. At least from time to time von Wright seems to think that neuroscience may have notable results for the autonomy of the mental. As I have shown, he thought that the work on the mirror-neurons of monkeys may be an advance towards finding type–type identities, and sometimes sees his own work as having an affiliation with eliminative materialism. Von Wright also noted that the kinds of cases where B (A) precedes M may show that an actionist description of what happens may seem to be “out of place.” Already in 1966 he had concluded that perhaps there will be a time when all action would appear to us as mere movement and would be described as such. He thus concluded that it was possible that causal analysis of behavior would replace the analysis of behavior in mental terms. Von Wright notes that from our current point of view, such a form of life perhaps looks to us stranger than the life of a Martian; it is nevertheless possible that this will be our fate.

1451 Von Wright, UPf, 5, my emphasis.
1453 Von Wright, UPf, 5.
When von Wright refers to neuroscientific evidence, he makes much use of the cases where B could occur before M. Let us nevertheless suppose that M occurs little before B. Is the causal efficacy of M then secured? No, because referring to Libet’s work we can say that the causal role of N is “already on its way” and therefore M, although perhaps part of N, cannot have a role in the causation of B. If voluntary movements are preceded by a certain kind of activation in the brain, N, which occurs before people make the decision to act, this could be a reason to say that although the decision M occurs before B, M is nevertheless predated by N, which without doubt is the cause of B. As von Wright says, referring again to the neuroscientific evidence, there might be reasons to think that N would have produced B also in the absence of M, which means that we should attribute causal priority to N in relation to M. Closely related to this is von Wright’s claim that if M and N occur simultaneously, then there cannot be a causal relationship between them. They are a joint effect of some physical event P (S in our diagrams), and we may have reasons to think that P could have produced N without producing M. Sound waves entering the ear can produce neural events without producing sensations. Interestingly, von Wright’s arguments for epiphenomenalism depend greatly on the considerations of the temporal relationships between neural and mental events. When considering whether mental states are causally efficacious it is crucially important to show their correct place on the time-axis, where causally efficacious neural events exist. This is a task which, to some extent, can be approached empirically as the neuroscientific research shows. But I think we should agree with von Wright’s conclusion that when it comes to timing of consciousness “[…] things are very tricky and unperspicuous here.”

If von Wright’s claim about the nature of the mental is taken seriously, we face a situation where consciousness would be causally cut off from everything. There must be a way out from this kind of understanding of consciousness if we value our view of ourselves as free agents. As far as von Wright’s position is concerned here, an important caveat must be noted. Despite the strong views that I have described, there are reasons to think that von Wright is not defending global epiphenomenalism. As he surprisingly says: “Of course, physical stimuli of sense-organs call forth sensations, and reasons move agents to act.” An important reason why Wright’s views do not imply global epiphenomenalism is his understanding of the concept of a cause. When von Wright says that consciousness stays outside the web of causally related events we need to understand what he means by a web of causally related events. He is talking about events which can be understood as being related in

---

1454 Von Wright, UPh, 158.
1455 Von Wright, 1999, 30.
terms of a relation which von Wright calls *Humean* or *nomic*. This relation satisfies the following conditions: 1) Cause and effect can be described in terms which make it possible to identify them independently of one another; 2) There is a law-connection between cause and effect. Both von Wright and Davidson emphasize that mental events cannot be related to physical events with this kind of relation. But this is not to deny that mental events can cause things and be caused by physical events in a looser sense of the word. We have seen that Davidson wanted to reserve room for a loose concept of cause. Stoutland has noted about von Wright’s position that:

In denying that explanations of behavior as intentional are causal… [von Wright] was denying that they are nomological, which does not rule out their being causal in some other sense that does not depend on nomological connections at any level…. 1456

I agree with this observation. Stoutland claims that when von Wright denied that reason is a cause he always had the Humean conception of cause in mind.1457 Von Wright’s understanding of the concept of cause may explain many of the confusions that can arise if the arguments for epiphenomenalism are read out of context. In the preface to *Explanation and Understanding* von Wright notes: “Those who think that actions have causes often use ‘cause’ in a much broader sense than I do when I deny this. Or they may understand ‘action’ differently. It may very well be, then, that ‘actions’ in their sense have ‘causes’ in my sense, or that ‘actions’ in my sense have ‘causes’ in theirs.”1458 Stoutland has noted that mere verbal differences are an important reason why on certain questions the positions of von Wright and Davidson have been misinterpreted as being completely opposite to each other. Perhaps von Wright’s epiphenomenalism can be clarified as well by focusing just on his use of terminology. I believe that Stoutland’s observation however neglects an important aspect of von Wright’s position, namely that the problem of epiphenomenalism in von Wright’s view does not relate *merely* to our understanding of the concepts of action, cause, or nomologicality. Stoutland’s claim that von Wright was not really disputing that mental explanations are causal in *some sense* ignores von Wright’s references to neuroscientific evidence. It is surprising that in the articles where Stoutland discusses von Wright’s views about philosophy of action or compatibilism he does not consider von Wright’s later position.

I believe that the conceptual reading cannot completely explain von Wright’s relation to epiphenomenalism. But there is one more – and very important – reason that von Wright’s

1456 Stoutland, 2005, 130.
1457 Stoutland, 2009. See also Stoutland, 2006.
1458 Von Wright, 1974, viii.
position can be seen to secure the reality of mental causation against epiphenomenalism. This is von Wright’s claim that the mental has epistemic priority over the neural. We have seen that at times von Wright argues that the mental is a “shadow” accompanying natural connections. The metaphor of mental being a shadow of the underlying physical processes is often used in formulating the position of epiphenomenalism. But, surprisingly, this is not von Wright’s last word. Because the mental has epistemic priority over the neural, it follows that the same priority holds also for rational explanations of action in relation to neural explanations of the movements involved in the acting. Although von Wright has suggested that the mental should be seen as having a shadowy role, his final conclusion, quite astonishingly, turns the picture completely around. He wants to challenge the modern view which, by the use of supervenience, lets the mental be the shadow. That contemporary philosophy of mind ascribes to the mental the same shadowy role as traditional epiphenomenalism is von Wright’s interpretation of the situation. This interpretation is not entirely out of place if we consider how eliminative materialism, for instance, understands the status of the mental. But when the picture is turned around, von Wright writes:

I think that...’neural epiphenomenalism’, as I propose to call it, is wrong. I would challenge it and reverse the perspective completely. I shall maintain that the neural chain is a ‘shadow’ in relation to the chain of mental events and that rational explanations of action have epistemic priority in relation to behavioural and neural explanations.

But how should we understand von Wright’s claim that actually the “neural chain is supervenient on the mental one”? In the standard understanding of supervenience, the relation is thought to be a relation of determination or dependence. Kim has argued that supervenience is best understood as a reductive relation. How does von Wright’s claim fit with this understanding about the nature of supervenience? It seems to me that we must conclude that it cannot be made to fit with the standard understanding of supervenience. The only way to make sense of the claim that the neural level is supervenient on the mental is to accept the epistemic priority of the mental and grant primacy to mental explanations as I have suggested. The conclusion of von Wright is:

If my argument is correct, it has a remarkable consequence for the ‘place of the brain’ in the mind–body problem-complex. Since the mental has epistemic priority in relation to the neural, and similarly

---

1459 See for example, Kim, 1996, 129.
1460 Von Wright, UPb, see also von Wright, 1998a.
1461 Von Wright, UPa.
rational explanations in relation to (neurological) causal explanations, the mental not only cannot be ‘reduced’ to the neural, but the neural cannot be used to ‘explain’ mental phenomena.

In essential respects this is a similar position to Davidson’s. The mental is irreducible and inexplicable in physical terms. Davidson’s reaches the conclusion through the arguments for the irreducibility of the mental which we have considered in chapter three, whereas von Wright uses the epistemic priority of the mental to reach the same conclusion. Von Wright’s argument is elegant and suggestive. I see it as a very tempting answer to the problem of mental causation and pointing in the right direction as far as a solution to the new problem of mental causation is concerned. I suggest that we could be satisfied with this conclusion. In section 4.2.2 I will describe how this view can inspire a non-reductive solution to the new problem of mental causation. Von Wright’s suggestion that rational explanations have epistemic priority in relation to behavioral and neural explanations seems to be on the right track and could be related to the epistemological or deflationary approaches of Burge, Baker and others. I believe this would be one clear way to bring the views of von Wright into the middle of recent debates in the philosophy of mind.

Since one central aim of this work is to clarify the views of von Wright, I unfortunately cannot be completely satisfied even with this conclusion. The reason for this is that von Wright’s conclusion described above is reached before the empirical evidence in favor of epiphenomenalism is given a role in von Wright’s writings. “Neural epiphenomenalism” is rejected in “In Defense of Psychology”, which dates back to 1996. This is consistent with the fact that von Wright ascribes the shadowy role to mental phenomena in “Mind and Matter”, which was published in 1994. At that time von Wright seemed to defend neural epiphenomenalism; two years earlier he had claimed that empirical evidence corroborates his claim that the mind has no independent role whatsoever in producing changes in the material world.

The most surprising thing is that after having rejected neural epiphenomenalism in 1996, von Wright goes on in 1997 to describe his position as a form of emergentism or epiphenomenalism and claims that consciousness stays outside the spatio-temporal web of causally related events. As noted already, von Wright’s views are not entirely consistent. In section 3.2 we saw that, according to von Wright, the fact that humans have brains can be ignored when considering the philosophical mind–body problem and related perplexities. Whereas in 1992 von Wright denies the importance of brain in the mind–body problem, in 1999 he claims that: “Serious thinking about [psycho-physical interaction]

\[\text{Von Wright, UPf, 4.}\]

\[\text{See von Wright, UPf.}\]
must take into account the role played by the nervous system (the brain)." I find it implausible that von Wright would have changed his views so dramatically about these important questions in such a short timeframe. What should therefore be our final conclusion about von Wright’s position?

Perhaps the final conclusion should not be forced. In the preface of *In the Shadow of Descartes* von Wright writes: “Some things may strike the reader… as mildly contradictory. It was not always possible for me to make up my mind definitely on alternative positions. I have not wanted to conceal or smooth out the agonies which thinking about the fundamental questions of philosophy always caused me.” If this is how von Wright saw his own position how could we smooth out the agonies? Although von Wright was willing to admit that a time might come when we stop to describe human life in terms of actions, he nevertheless wanted to vigorously defend, as the epistemic priority of mental and rational explanations show, the autonomy of agency. The possibility entertained by von Wright of future humans who see each other not as acting but merely moving is nevertheless interesting.

As we saw in section 3.2, at one point von Wright claimed that actions could perhaps be read off from the brain. Davidson and Quine, both anomalous monists, claimed that such a reduction to neurology cannot happen. Quine, despite his eliminativist views, admitted – in a very comforting spirit to my ears – that: “It might be, now and forever, that the only way of guessing whether a man is inspired or depressed, or deluded, or in pain, is by asking him or by observing his gross behavior; not by examining his nervous system.” This is what was suggested in the laboratory example of section 3.2. Kim, often considered a modern reductivist par excellence, admits that mind cannot be completely reduced to the physical and that the behavioral level is indispensable. It is interesting to note that Quine and Kim, for example, seem to stress the importance of behavior whereas von Wright at times is willing to make the more daring claim that mental phenomena could be reduced to neural phenomena. As a curiosity we can further note that in the turbulent years for the philosophy of mind in the 1960s von Wright was willing to defend epiphenomenalism whereas Davidson defended the choice to see human beings as an anomaly. I conclude, however, that von Wright dismissed these eliminativist views at the end of his career. This being said, these views perhaps suggest

---

1464 Von Wright, 1999, 29. To what kind of serious thinking is von Wright referring? Is he referring to serious philosophical thinking or to some other form of serious thinking? From the von Wrightian perspective this distinction is possible to make. But does serious philosophical thinking about the mind need to concern itself with the role played by the brain?

1465 Von Wright, 1998, xi. From the notes attached to the unpublished papers it is possible to see that the “agonies” continued until the end of von Wright’s philosophical journey.

1466 Quine, 1976a, 243.
that the conflict resulting from trying to combine the teachings of Kaila and Wittgenstein was something which von Wright never managed to solve completely.

Without suggesting that this is von Wright’s view I propose that, as long as we want to see ourselves as free agents, we cannot really allow mental events to be connected to physical events by a nomic relation. In this sense consciousness, as long as freedom is an essential feature of it, cannot be part of the spatio-temporal web of law-governed events. It is useful to recall Davidson’s famous claim: “The anomalism [lawlessness] of the mental is… a necessary condition for viewing action as autonomous.”

I think that von Wright makes a similar point with his dramatic claim that causation is restricted to physical phenomena of nature and that consciousness has no place there. Von Wright once noted that it is not possible to raise one’s arm and observe the cause as well. A Humean cause and a reason for action cannot thus co-exist at the epistemological level. The claim is that:

When I observe, I let things happen. When I act, I make them happen. It is a contradiction in terms both to let and to make the same thing happen on the same occasion. Therefore no man can observe the causes of the result of his own basic actions.

Davidson makes a similar point: “When we are making up our minds what to do or what to think we cannot at the same time conceive of our reasoning as bound by the strict laws of physics.” This is a partial answer to the problem of how reasons can explain even though they are merely supervenient on the efficacious physical causes. The epistemic practice of explaining actions in terms of reasons, as long as it works and as long as we keep on using it, must be seen as carrying a certain ontological import. Yet, there remains a puzzle. How can we be free if everything is strictly determined? Is the view that we are free, that mental causation is real, only a result of our ignorance about the neurological causes of actions? Already in Explanation and Understanding von Wright notes: “The conceptual basis of action, one could say, is partly our ignorance (unawareness) of the operation of causes….”

This can be compared to Davidson’s observation that “Too many of the causes and effects of human thoughts are unknown to human thought.” It could be said that we believe in mental causation because we do not know the underlying neural mechanisms and we cannot reach them because no man can observe the causes of his own basic actions. This motivates

1467 Davidson, 1970, 225.
1468 Von Wright, 1974, 130.
1469 Davidson, 1999u, 359.
1470 Von Wright, 1970, 30.
1471 Davidson, 1993f, 312.
the need for reason-based explanations. The view expressed in *Explanation and Understanding* is repeated thirty years later when von Wright concludes:

> No assumption about ‘gaps’ in the causal order is needed in order to understand the possibility of action. The result of an action is caused (Humean) by muscular movements which in their turns are caused by impulses from the brain to the muscles. But the cause (Humean or not)...from the brain to the muscles we simply do not know...there may be such causes or there may be not.  

This insight expressed in an unpublished note tellingly titled “Eureka!” shows that in von Wright’s final view, it is our unawareness of the Humean causes that makes action possible. It could be argued that it is only our incomplete or inadequate knowledge which makes mental causation possible. On the other hand, our belief in agency is a result of those experiences which indicate to us that we can make things happen. In *Explanation and Understanding* von Wright describes the other part of the conceptual basis of action which is: “[...] our confidence that certain changes will happen only when we happen to be acting.”  

Also this aspect is repeated when von Wright states his final conclusion: “[...] even a convinced determinist...would say that we act and agree that a change produced by as agents would not have occurred had it not been for the interference of the agent.”  

It could be said that the confidence that we can act is epistemic, but the fact that we have it has an ontic foundation in certain regularities in the world. Incomplete knowledge about (mindless) nature together with our complete knowledge about our own nature opens the space for mental causation.  

Empirical evidence, such as the kind of neuroscientific findings that we have discussed, suggests that mental explanations may not be causal in any sense. I find it interesting that von Wright was so eager to stress the relevance of neuroscientific evidence as corroborating his thesis about the mind’s causal inefficacy. It seems to me that the views of von Wright as described in section 4.2.1.1 do not fit well with his admittance that “of course” reasons move agents to act. Von Wright’s repeated references to epiphenomenalism and his formulation of the mental–physical relation in the way characteristic of traditional epiphenomenalism raises the question of how von Wright can also pledge the causal efficacy of the mental. What can be said at the very least is that von Wright’s references to epiphenomenalism are unfortunate if they are not meant to implicate some kind of real causal inertness of the mental. If references to epiphenomenalism allow mental causes to be the kind of causes that Stoutland suggests, then von Wright chose his words poorly. I believe von

---

1472 Von Wright, UPI.
1474 Von Wright, UPI.
Wright did choose his words carefully and therefore I take the references to neuroscientific evidence as suggesting that he really wanted to deny the causal efficacy of certain mental phenomena, or at least wanted to allow the possibility that these phenomena may turn out to be without causal efficacy.

4.2.1.2 Apparent mental causation

We have seen von Wright’s two reasons, conceptual convenience and empirical evidence, for defending a version of epiphenomenalism. We could accept von Wrightian epiphenomenalism as a “negative solution” to the problem of mental causation. Although the possibility of epiphenomenalism cannot be refuted, few would be willing to rest with the conclusion that consciousness stays outside the spatio-temporal web of causally related events or that sensations are neither caused nor causally efficacious. Only few philosophers are willing to rest with the epiphenomenalist conclusion if the epiphenomenalist claims like von Wright’s are meant to indicate real causal inefficacy.

But if mental causation is unreal, if mental states and events are really mere epiphenomena as von Wright suggests, then why do we so often think differently? On what is based our conviction that mental states and events are causally efficacious? I believe that when thinking about this aspect of the problem of mental causation, we can benefit from von Wright’s observation that the relation of the mental and the physical is semantic. Perhaps this helps us to understand why we so often think that there was a relation of the form “mental cause – physical effect”. It seems to me that we must offer some explanation for why we so often see our behavior in terms of this relation. But to say that we need to give this kind of explanation is of course not to say that there actually is such a mental–physical relation involved in the cases which we explain in terms of it. One possibility, somewhat inspired by von Wright, to think of the “mental cause – physical effect” relation could be the following.

The diagram describes a situation where a physical event P in the brain causes behavior. The same physical event P causes also a mental event M, in this case a conscious experience. This
may create the *illusion* that M is the cause of B. From this perspective, behavior B would be understood as action A having M as its reason-cause. The arrow between B and A is meant to indicate that, at the level of behavioral manifestations B and A are identical. But although there is a causal relation between P and B as well as P and M, it *could* be the case that there is no causal relation from M to A – which would mean that although M temporally precedes B / A, there is no causal relation between them.

Some psychologists have considered the cases where we seem to think that our mental states were causes of our actions in the way mentioned above. I am referring to a model suggested by psychologists Wegner and Wheatley.\(^{1475}\) What they call “a model of conscious will” looks like this.

```
Experience of conscious will

Thought                   Apparent causal path         Action

Actual causal path

Unconscious cause of thought       Actual causal path

Unconscious path

Unconscious cause of action
```

Wegner and Wheatley claim that people can experience conscious will quite independently of any actual causal connection between their thoughts and actions. This claim they take to be fairly uncontroversial from the perspective of empirical findings.\(^{1476}\) The model of conscious will described above is meant to be a model that is consistent with these findings. In this model, conscious will is experienced to the degree that an *apparent* causal path is *inferred*


\(^{1476}\) They refer, for example, to the findings of Wilson and Nisbett (1977) and to the research by Libet (1985) which attempts to show that voluntary action is always preceded by an unconscious cerebral initiative. For more references to similar studies see Wegner and Wheatley, 1999 and Wegner, 2003a. Wegner and Wheatley (1999) describe also their own experiment which is meant to show that people can be led to experience willful action when in fact they have done nothing. The experiment is too complicated to be described here in its entirety and therefore the interested reader should view the original article.
from thought to action. The better the chances for such inference, the clearer the feeling that one is “in charge”.

The model describes a flow of various events leading to a voluntary action. On the one hand there is an unconscious cause of action. We see an actual causal path leading from the unconscious cause to the action. This can be understood as a physical cause. On the other hand, there is an unconscious mental process which gives rise to conscious thought.\footnote{I hesitate to use the expression unconscious \textit{mental} process. I would thus liken this process to a physical process of which we are not aware.} There may be a link between these unconscious processes, but there need not be. The question of whether there is such a link is irrelevant to the perception of the \textit{apparent} path from conscious thought to action. Wegner and Wheatley conclude: “There need be no actual path here, as it is the \textit{perception} of the apparent path that gives rise to the experience of will: When we think that our conscious intention has caused the voluntary action that we find ourselves doing, we feel a sense of will. We have willfully done the act.”\footnote{Wegner and Wheatley, 1999, 482, my emphasis.} The problem is here formulated in terms of free will, but the implications for the problem of mental causation are obvious. In the diagram the arrow “experience of conscious will” could as well read “experience of mental causation.” The simple explanation of Wegner and Wheatley is that: “[…] it may be that people experience conscious will when they interpret their own thought as the cause of action.”\footnote{Wegner and Wheatley, 1999, 480.} I think that the claim that apparent mental causation is a result of our incorrect self-interpretation is an interesting and plausible suggestion. According to Wegner and Wheatley the source of this misinterpretation could be:

\[\text{[\ldots] a causal illusion that is the psychological equivalent of the third variable problem in causal analysis. We can never be sure that A causes B, as there could always be a third variable, C, that causes both of them. In the same sense, we can never be sure that our thoughts cause our actions, as there could always be unconscious causes that have produced them both. The impression that a thought has caused an action rests on a causal inference that is always open to question.}\] \footnote{Wegner and Wheatley, 1999, 482.}

This sounds very plausible and relates to the general problem which I raised earlier. How can we be sure that there is a causal relation between the events that we perceive? It seems that when it comes to a specific case of mental causation there is nothing \textit{in the experience of mental causation itself} that could guarantee that our thoughts cause our actions. What could convince us that a third variable has not interfered with the chain existing between our thoughts and actions? We have the ordinary experiences which give strong support for the
view that mental causation is not apparent, but unfortunately these experiences do not show that we could not be mistaken in each individual case. Although, for the sake of sanity we should not claim that global epiphenomenalism is true, we have to accept the possibility that in many cases where we seem to observe mental causation the phenomenon of mental causation is only apparent. Evidence from psychology supports this conclusion. When philosophers consider the problem of mental causation they usually start from the assumption that seemingly clear cases of mental causation are cases of mental causation – and then the problem is to explain how this kind of phenomenon is possible in a physical reality. But if we do not have a fool-proof method for distinguishing the apparent cases from the actual cases, then the prospects for a metaphysical solution to the problem of mental causation look dim.

Apparent mental causation is a result of our incorrect self-interpretation. But in a similar fashion, we could claim that “real mental causation” is a result of our correct self-interpretation. My tentative suggestion is that the problem of mental causation is solved when we accept that mental explanations are the best evidence of the occurrence of the relation of mental causality. The question of how the causal relation itself is possible is irrelevant. If the mental is not an ontological category in any clear sense, it is perhaps being too optimistic to wait for an ontological solution to the problem of mental causation. Perhaps we should, after all, agree with von Wright that there isn’t substantial relation of “mental cause – material effect” or vice versa. The idea of this relation is a result of a certain kind of interpretation we make of ourselves. The facts on which this interpretation is based are numerous; there are various psychological factors, the existence of which strengthens an agent’s conviction that he was the cause. Wegner and Wheatley claim: “Because we have thoughts of what we will do, we can develop causal theories relating those thoughts to our actions on the basis of priority, consistency, and exclusivity.” This suggestion can be compared to the recent suggestion of neuroscientists: “The subjective (and potentially illusory) feeling that we are executing a movement does not arise from movement itself, but is generated by prior conscious intention and its predicted consequences.” The priority, consistency and exclusivity are important factors in person’s self-image as a being with a conscious will. When we see a connection between thought and action, we expect that the thought should occur before the action

1481 We can still understand that there is a desire to ask this how-question. But the only way to even to try to answer it would require a quite concrete reduction of the mental to the physical, and all the irreducibility arguments so far suggest that this project is hopeless. As noted in section 4.1 we can of course accept the view that the problem of mental causation is solved when we state that mental phenomena are physical phenomena and the latter are causally efficacious. But is this an answer to the problem of mental causation?
1483 Desmurget et al., 2009, my emphasis.
(priority), be consistent with the action (consistency), and not be accompanied by other potential causes (exclusivity).

These three principles seem to relate to our more general understanding of the nature of causality. This understanding is something which we have learned. It seems, for example, that our understanding of the exclusivity principle depends on our learning of what kind of things can be causes of what kind of other things. There is evidence showing that humans tend to discount the causal influence of X on Y if there is potential Z that could be also a cause. Whether Z can be a causal factor on this occasion depends on our general knowledge and understanding of how reality works. When applied to the case of mental causation we could perhaps say that what kind of mental causes a person can detect is partly dependent on what kind of psychological understanding a person has developed. According to von Wright, the capacity to have reasons for one’s actions develops when a child “[…] grows up to be a member of society, learns to speak and do various things, to understand the meaning of challenges and institutions, and to participate in various practices….”  

Once this “form of life” is learned, the capability of seeing semantic relations between behavior and mental phenomena becomes possible. Interestingly, Wegner and Wheatley note that “the principle of consistency in the experience of will draws on the observation that the thoughts that serve as potential causes of actions typically have semantic associations with the actions.” The semantic associations to which psychologists refer could be linked to von Wright’s more philosophical observation that the relation between the mental and the behavioral is semantic.

It could be claimed that the proposal that real mental causation is a result of our correct self-interpretation immediately faces an obvious problem, namely the possibility that people interpret themselves incorrectly. If people infer wrongly the relation of mental causation when there is none, how could self-interpretation serve as a solution to the new problem of mental causation? I admit that this is a severe problem, but I nevertheless suggest that, at least in the absence of conflicting evidence, there is a presumption that a person acted because of a mental cause. Wegner and Wheatley note: “The experience of will can be an indication that mind is causing action, especially if the person is a good self-interpreter, but it is not conclusive.”  

I wholeheartedly agree with this view, except for the last sentence which I recognize to be exactly the point that the psychologists want to make. It seems to me that an experience of will is surely a good indicator of a mind causing an action, and the more

sensitive a man is to the workings of his mind, the more aware he is of himself, and the better the chance he has to observe whether he is really the cause. We can agree that the experience is not conclusive, but I think we have to agree that it is hard to think of evidence which would be more conclusive. This is where the discussion has to end, since lacking mental–physical reductions we have no other way to approach the problem.

One could say that how we interpret the agent’s view about his role in the causation of his actions is related to the way we see, or want to see, the whole situation. If we think that he was the cause of what happened, we are taking him to be responsible for the obvious consequences, i.e. to be a candidate for moral evaluation; in short we take him to be an autonomous human being. Perhaps the agent’s own certainty could be measured by considering to what extent he is committed, for example, to the consequences. The presumption that a sincere agent is the cause of his actions does not give us absolute certainty about what happened, but we cannot do any better. If we consider the neuroscientific and psychological evidence, we have reasons to believe that the mental phenomena in terms which we explain our behavior are epiphenomenal from time to time. I suggest that we should therefore conclude that “conditional epiphenomenalism” is true. It is good to realize this, because it makes us consider the extent to which we are, or think we are, transparent to ourselves. How often do we know the reasons which influence us and how often do we make false inferences and thus summon up the phenomenon of apparent mental causation? The possibility of conditional epiphenomenalism reminds us how poorly we know ourselves.

In addition to what has been already suggested, I think there are two general observations which could be used to defend the view that we are mysteries to ourselves and that it is not clear when or how often we in fact make conscious decisions. On the one hand, let us consider the simple event of the occurrence of a certain thought. Suppose that you are lying on your back, watching the sky, when suddenly a thought that you should call your brother occurs. Or suppose that I instruct you to “clear your mind” and to think nothing. It is not easy to maintain a completely blank mind; it is actually extremely difficult especially if instructed to do so. Sooner rather than later some thought occurs in your mind. You cannot control the thought, you cannot prevent it from occurring and you cannot decide what kind of new thoughts will occur as a result of the first thought. In this sense you are really a victim of your circumstances. Why a certain thought occurs literally from nowhere is a truly puzzling, even disturbing question. I think we could say that this is real emergence, the case of neural

---

1487 For further neuroscientific considerations, see Pockett, Banks, and (eds.), 2006.
1488 For examples and for a very interesting discussion see, Wilson, 2002.
“quantity” turning into mental quality. Even if we could demonstrate that a specific thought was “realized” because neural state N occurred, would this be a satisfying answer to the question of why this unique thought with the specific content C occurred? Why did the neural state realize just this thought? In the end, the thoughts we happen to have are not up to us and neither can we control the moment when they occur.

Closely related to this realization is the insight suggested, for example, by Galen Strawson. It is old wisdom, pointed out by Schopenhauer among others, that a man can do what he wants to do but he cannot determine what he wants. Whereas we sometimes can control what we want by desperately pushing our desires towards a certain direction, there is always the troubling question of where, if anywhere, the original desires come from. They seem to be just a result of the fact that we are the kind of persons we are. Environment and heredity entirely shape our characters. But we are not, and cannot be, responsible for our environment and we certainly cannot be responsible for our heredity. So it could be claimed that in a fundamental sense we are not, and cannot ever be, responsible for our characters, and this being the case, we cannot be responsible for what we do. Whatever I do is determined by my past self and I had no control over the events that shaped me in the past. This self-determination is something whose consequences for our capability to choose freely should be considered more carefully. Such a task cannot be carried out in this work, but it seems to me that it would support the case of conditional epiphenomenalism or at least clarify the question of how poorly we know ourselves and how little control we have over our decisions.

Whereas conditional epiphenomenalism looks to me to be very plausible, global epiphenomenalism is more problematic. We have no good reasons to believe that we would be mistaken about ourselves all the time. If we consider how well our explanations work, it would be irrational to claim that they always fail to capture the mental causes behind our actions. If mental explanations are really always pseudo-explanations then, to use Fodor’s words, that simply is the end of the world.

4.2.2 Precursory reasons

In this section I suggest a way to think about the new problem of mental causation. As noted at the beginning of this chapter, the approach that I find most congenial can be described as an epistemological or deflationary approach to mental causation. It is partly inspired by the “non-

1489 Strawson, 1986.
causalist” aspects of von Wright’s and Davidson’s positions and it partly depends on the distinction which was briefly suggested in section 2.5.1 between beliefs and beliefs.

I have repeatedly claimed that a non-reductive position cannot provide a reductive explanation for the problem of mental causation. My suggestion is that if, as von Wright and Davidson claim, there is no scientific solution to the problem of mental causation then the real problem for any non-reductive view about the mind is to answer the question of how to decide when an explanation which refers to mental causes is true. This question is relevant for our self-conception as agents that can act out of reasons, choose freely their actions, and in this sense influence the course of their lives and shape their destiny. If we cannot tell when our explanations of actions are correct, we face a situation where our conception of agency is constantly jeopardized. If we cannot explain when mental explanations of behavior are correct we have to, or least can in principle, doubt our reasons all the time and speculate about whether or not we are in charge of our own actions.

The question of “when I am in charge and how to decide this” is related to pragmatic problems, such as: when can a person be held responsible for her actions? If it cannot be shown that a mental cause was operative in the production of an action then a person could, in principle, claim that it was not (s)he who did something shameful or blameworthy, it was her/his brain. In the cases when we conclude that the person was not in charge of their actions, they cannot be judged according to the same standards as “normal” people. Sometimes we accept that the person as a deliberating and autonomous individual could not prevent what their body caused.\textsuperscript{1490} The question of how to solve the new problem of mental causation is interesting from a philosophical point of view, but it has also relevance for the question of when a person is responsible for what they did. I have concluded that the impossibility of a reductive explanation for the problem of mental causation leads to the possibility of conditional epiphenomenalism. A solution to the new problem of mental causation does not remove this possibility; it merely states the conditions which must be satisfied in order to solve the new problem. My suggestion is that the possibility of conditional epiphenomenalism is something which remains even if we provide a solution to the new problem of mental causation. Conditional epiphenomenalism is thus an important part of the answer to the traditional problem of mental causation.

The new problem of mental causation is the problem of how to tell when a mental explanation is true. When does this kind of explanation refer to something which is causally

\textsuperscript{1490} For a discussion. see for example, Glannon (ed.), 2007, Gazzaniga, 2005.
effective in the production of behavior? This question cannot be answered by claiming that such an explanation is true when mental causes are identical with the neural causes resulting in behavior. Whereas this claim is what the following answer implies, it cannot be used as a ground for the truth of mental explanations because it cannot be shown that there actually was a neural cause corresponding to the mental cause in terms of which an action is explained. Our inability to show the correspondence between the causes leads to the possibility of conditional epiphenomenalism. If we now try to solve the new problem of mental causation, we have to consider how to decide whether an explanation like “Georg bought apples because he likes their taste” is true. I will consider this question by describing the famous pantyhose experiment of Nisbett and Wilson.\footnote{Nisbett and Wilson, 1977.} In this experiment Nisbett and Wilson set up a market survey table outside a shopping centre, placed pairs of pantyhose on the table and asked the random shopping centre customers which of the various pairs they preferred, and why.

Suppose, in the spirit of the experiment, that there are four pairs of pantyhose A, B, C, and D in front of me. I am asked to choose one and I choose pair D by picking it up. When asked why I chose that particular pair I answer: “Because the material looked good.” Then we have the following situation.

\[ \text{The material looked good (R)} \rightarrow \text{Picking up pair D (A)} \]

In a certain situation I have a reason (R) for action, an action (A) occurs and we think there is a relation between (R) and (A). We can call this relation a causal relation, but should understand it in the spirit of what we have learned from von Wright. It is not a nomic relation, but causal in some other sense.

In the original experiment all pairs of pantyhose were identical and this fact was not known by the participants. Nisbett and Wilson, on the other hand, knew in advance that given a choice among three closely-matched alternatives, there is a bias to pick the last one. The human cognitive system is a curious apparatus and many facts of how it “makes choices” or “interprets” data below the conscious level are known. It is known, for example, that people tend to prefer objects on their right and therefore we can predict that, given identical pairs, pair D will be over-chosen. Indeed, in Nisbett’s and Wilson’s experiment pair D got picked up almost four times as often as pair A. The rightmost pair was heavily over-chosen, but the test-subjects did not give the position of the pair as their reason for picking it up. This being the
case, there is a good reason to think that the people who chose pair D did it just because the pair was the furthest one on their right side. Then the situation is like this.

\[
\text{Pair D was on the right side (C) } \rightarrow \text{ Picking up pair D (B)}
\]

Here we have a cause (C), behavior (B) and a causal relation between them. Reason does not belong to this situation and therefore, according to the view that I have been trying to put forward, the picking up of pair D ceases to be an action (A). Let us suppose that so far the situation is as described. I pick up pair D and give my reason when asked. I have thus given a reason-based explanation. People responsible for the experiment think that I picked up pair D because it was on my right side, and they give a causal explanation of what has happened. The question that I want to raise is: How to decide which explanation is correct? Both cannot be correct; if I picked up pair D because the material looked good it was not the case that I picked up the pair because it was on the right side. If, on the other hand, I picked up the rightmost pair because it was the rightmost, then I did not pick up pair D because its material looked good. If we believe that the mental will not be reduced to the neural, then we cannot identify my reason with a cause; an answer which argues that a neural state caused my picking (and we know the real cause when we identify the causally efficacious neural state) cannot help. Moreover, the explanation referring to C refers to the human visual system and its physical properties, whereas the explanation referring to my reason does not refer to such properties. It could therefore be the case that my visual system is working normally and unconscious processing is taking place in the brain simultaneously with the neural state, which would be understood as my reason if the identification made sense. There would be thus two candidates for a neural cause. Which one of these would be the real cause?

Let us suppose that after my choice I am immediately asked why I chose pair D and I answer: because it had the smoothest material. We could doubt this explanation because all the pantyhose were identical; yet we cannot deny that it could have been my real reason; maybe, for some reason or other, I really felt that pair D had smoother material than other pairs. To insist that I could not have felt this because the pairs were identical is an absurd claim. The person in charge of the experiment is referring to his knowledge and claiming authority over the explanation, and I do the same. Who is correct? The question is more severe in cases where the issues at stake are more serious. We can assume that knowledge about the brain could be used in cases where it is important to decide whether or not an agent is lying about his actual reasons. In these cases a conflict between the agent’s own view and
the view of an outsider could have drastic consequences. The interest in brain-scanners or mind-reading machines is clearly increasing, and there are already various suggestions about how such new technology could be used.

Not surprisingly, institutions like law-enforcement agencies and the military have expressed a lot of interest in mind-reading technology. The Churchlands, just to use one example, have speculated about how neural facts could be used to draw the line between normal and abnormal people. I believe we are aware of the ways in which something like this has been happening for some time. Mental illnesses are being diagnosed in terms of physical evidence, and medicines are offered instead of therapy. Whereas currently it seems to be trendy to claim that mental illnesses, addictions and other problematic mental–physical phenomena need to be understood from a perspective which draws together the results of psychology, brain-sciences and perhaps even sociology and philosophy, we can imagine a situation where the neural evidence alone would become the criterion for these problematic phenomena. It seems to me that a philosophical view, like the one put forward by neurophilosophers, is simultaneously a suggestion about the nature of the desirable society which these philosophers would like to pursue. Let us suppose, referring again to the imagined experiment, that I am not aware of my reason for picking up pair D but that I can state it when challenged.

We can play with counterfactuals and say that if this really was my reason then I would have been able to state it before I acted. But this observation is of no use if I actually was unaware of the reason before the action. What can guarantee that the reason that I give is not an invention? In Nisbett’s and Wilson’s experiment, people gave all kinds of reasons for picking up D but all these reasons, if based on the differences between the pair of stockings, look suspicious because there were no differences between the pairs. People offered fabricated reasons which, given the situation, sounded reasonable to them. It seems that people tried to answer intelligently, and in the original experiment no one offered as a reason the position of pair D. In fact, when people were offered this explanation, they treated it as being very strange. On the view that I am proposing this is no surprise, because this fact was actually not a reason, it was a cause, and we have no reason to think that humans are usually aware of the causes of their behavior. The mechanism responsible for the fact that people tend favor objects on their right side may very well be hidden from our conscious access; likewise for many similar kinds of mechanisms.

My claim is that if a person picks up pair D because it is on his right side and gives a fabricated reason, then the picking is not an action. What happened can be explained in
physical terms, in terms of a causal explanation. It is nevertheless plausible that people firmly claim that they had a reason for what they did and they are “absolutely sure” that they did not invent this afterwards. It seems to me that if we do not require the awareness of one’s own reasons before the action, we cannot tell whether the picking of pair D was an action. In this case, we should trust the person responsible for the experiment and conclude that I picked up pair D because it was on my right side. D’s position caused “me” to choose it. But what should we say about the situation if I was aware of my reason before the behavior?

I suggest that a crucial distinction between precursory and retrospective reasons needs to be made. Precursory reasons are those of which an agent is aware before the relevant action is carried out. Retrospective reasons are those to which he refers after the action has been performed and of which he was not aware before the action. I can give the reason for my action in two different ways, namely before the action takes place or after the action has already been performed. I believe that a distinction between two kinds of reasons is required in order to see the difference between these two kinds of explanations. Let us suppose that I was not aware of my reason for picking up pair D before my behavior and before the question about my reason was put. Yet, immediately when I was asked about my reasons, I gave the reason “the material looked good”. In the terminology that I am suggesting, I would be giving a retrospective reason, which is something that becomes visible only after the action as a result of self-observation. The essential question is: what is the explanatory strength of a retrospective reason in the pantyhose scenario? What is the guarantee that this kind of rationalization of action is not my invention? If a precursory reason is mere invention the existence of which is confirmed after an action has taken place, what would be our reason to think that such a reason was a causal factor in what happened? Is the reason relevant for the fact that I picked up pair D? If the reason did not exist before the action, how could it have causal relevance for the action that occurred? The problem with the explanation that refers to a retrospective reason is that the temporal order of the reason and action seems to be such that an explanation of the latter in terms of the former is not possible.

My claim is that if I explain why I chose pair D by referring to a retrospective reason, then we are entitled to think that outsider’s explanation of the behavior by referring to the position of D could be correct and my sincere explanation false. This would mean that there is no mental explanation for “my” behaviour, and an outsider would be the supreme judge with respect to the question of what happened. It is a statistical fact that people tend to prefer objects on their right side. This, together with the fact that we do not have a satisfying mental explanation of why I picked up pair D, is a reason to think that a non-mental explanation is
satisfying. Lacking a better answer and knowing that in situations like the one described in the pantyhose example people are often affected by the position of the object, a non-mental explanation is *reasonable*. The cause would override the reason, and the outsider would override agent’s autonomy and self-knowledge.

The situation changes in a crucial way if I have a precursory reason for my picking. Suppose that before I make my choice I am aware of why I am choosing pair D; simply because the material looks good. We could debate whether it is possible that one pair looks better than the other if the pairs are identical. This, whether I can prefer one over the other if all are identical, is a good question. Let us, for the sake of the argument, fabricate the following story. I examine the pairs and when the time of pair D comes, for some obscure reason or another, a pleasant sensation or thought occurs in me. Unbeknown to me, this pleasant sensation makes me to see pair D in a “fashionable light” and a thought “This material looks good” occurs to me. This is surely something that *can* happen. Then we have a situation like this.

\[
\text{S / T} \rightarrow \text{PR} \rightarrow \text{A}
\]

A pleasant sensation (S) or thought (T) *causes* a precursory reason (PR), which is a reason for my action (A). *Why* PR exists does not really matter, what is important is that it, *de facto*, occurs. The case can be made without imagining the intervening pleasant sensation. I pick up pair D because of my precursory reason that the material looks good. It is my belief that the material looks good, which causes me to choose the pair, but the belief may of course be false. Nevertheless, it truly was my reason for picking up D.

If I am aware of my reason before I make my choice, there is every reason to believe that this reason is what made me act. The situation is a mirror-image of the situation where I give a retrospective reason. In the latter case, it is the explanation of the outsiders that we should accept, whereas in the former case we cannot sidestep the fact that I *actually had* a reason for my action. When I have a precursory reason, the fact that people tend to choose objects on their right side becomes secondary – as far as explanation is concerned – in comparison to the reason I had. If I have a reason for my action and I know this, how could I accept the claim of an outsider that this reason was in fact not my reason for my action? The fact that people tend to choose things on their right side and the fact that I chose pair D does not exclude the possibility that I did it for a reason. The lesson is this: in the case of
retrospective reason, we have no ground for saying that an action was performed whereas in
the case of precursory reason we have no ground to say that an action has not been performed.

A precursory reason we can understand as a thought of which I am aware before the
action. A precursory reason comes close to von Wright’s definition of “Cartesian
consciousness.” A subject is in this sense conscious if he is, for any given mental state M,
aware of whether or not he experiences M or not. A precursory reason is something of which
the agent is necessarily aware, because having a precursory reason simply is a conscious
reflective thought. It is for this reason that I have emphasized that an essential distinction
between beliefs and beliefs needs to be made. Only beliefs can function as precursory
reasons and only precursory reasons can secure mental causation. From an ontological point
of view we can ask what the status of beliefs is. We are aware of beliefs. In the context of
O-physicalism, I find it difficult to see what else beliefs could be than internal entities
“running through the subject’s mind”. We can recall Child’s analysis of the central lesson of
interpretationism: "The attitudes we cite in explaining an action need not correspond to
anything running through the agent’s head before she acted.” This is true, but then we
should ask whether the attitudes used in mental explanations really cause anything.

In the pantyhose example I would have a precursory reason if I had a conscious
thought “This material looks good” before I proceed to pick up pair D. Let us suppose that,
when asked, I would give a reason of which I was aware before the action occurred. In this
case, it would be difficult to see why or how a neurological or any other non-mental
explanation could override my sincere judgment. People tend to over-choose objects on their
right side, but what would be the relevance of this fact for the particular case where there
exists a mental explanation in terms of precursory reasons? In those cases where an agent
gives a retrospective reason, a non-mental explanation may have overriding potential with
respect to the agent’s own explanation because the non-mental explanation refers to
something which did exist before the action occurred, whereas the mental explanation refers
to no such thing. Something caused the action and in the absence of a precursory reason a
non-mental explanation may very well be acceptable. However, in the cases in which a
precursory reason does exist, there is no rationale for accepting the truth of non-mental
explanations. A precursory reason is a genuine mental cause and a non-mental explanation
cannot override an explanation which refers to such a cause. What could convince us that a

1492 Child, 1996, 123.
precursory reason was not the agent’s reason for acting? How could an outsider know better the inner life of a subject than the subject himself?

An explanation referring to a precursory reason thus has priority over an explanation which refers to a non-mental cause. I suggest that a non-mental explanations cannot override explanations referring to precursory reasons, and that an explanation referring to retrospective reasons does not automatically override a non-mental explanation. In the context of a non-reductive position, the problem of mental causation is re-formulated as the problem of how to tell when a mental explanation is true. My tentative answer to this question is that we can trust only those mental explanations which refer to precursory reasons. The model presented here is, of course, very crude and must be developed further in the future. I have offered only an outline of a view which to me looks far more reasonable than a metaphysical approach to the problem of mental causation; my purpose has been only to give a new way of thinking about the problem. The specific questions on which this model should focus in the future relate especially to the problems concerning self-deception and to the human desire to invent rationalizing explanations. Whereas these are interesting questions for psychology, clarifying the conceptual distinction between reasons and causes and between action and behavior is the task of philosophy, and empirical data alone cannot tell us how we should understand experiments like the one described by Nisbett and Wilson and the results of these experiments. I hope that the general idea, and above else motivation, behind the model which emphasizes the importance of retrospective reasons is clear.

I believe that the claim that we can trust only those mental explanations which refer to precursory reasons is in essential respects in harmony with von Wright’s view that the mental has epistemic priority over the neural. Whereas von Wright’s and Davidson’s views about the special nature of reason-based explanations and about the irreducibility of the mental have inspired the kind of model which I suggest here, it seems to me that their views face certain difficult problems with respect to the precursory / retrospective reason distinction and the distinction between beliefs and beliefs. By using the precursory / retrospective distinction, let us consider a more complicated example than the Nisbett–Wilson experiment. Let us suppose that I go to a party. Upon my arrival I am asked “Why did you come? It is rare to see you at parties.” I answer: “I wanted to see my friends.” We can assume that the person who asked the question is usually satisfied with this kind of an answer and so am I. The questioner would, in his philosophic mood, conclude that since I came to the party and gave the reason that I wanted to see my friends, I also believed that my friends would be present and these facts jointly caused me to come to the party. The answer I gave is a perfectly adequate answer.
to the question; it is understood immediately and we can “see” a connection between my answer and my action. One could say, as von Wright does, that here my action can be understood in the light of the reason without inferring a causal connection, i.e. the action becomes perfectly understandable when the reason is known, and this is the end of the story. But could we still ask the question of what my real reason for coming to the party was? Did I just come up with an adequate and reasonable answer which made sense when faced with the question “Why did you come?” This would mean that I would have given a retrospective reason for my behavior. Or had I been contemplating in silent von Wrightian monologue the question of whether I should go, and decided that I wanted to see my friends? This would mean that I would have had a precursory reason for my action.

Davidson would say that my desire to see my friends was, together with my belief, my reason for going to the party if it caused me to go. The reason which caused me to go is the reason why I am at the party. But to insist on this does not help us to answer the question of which of the reasons was actually my reason for going, or whether there was a reason at all. It seems to me that von Wright and Davidson do not make a clear distinction between precursory and retrospective reasons, which leads to the problems we have described.

Davison notes:

> We cannot suppose that whenever an agent acts intentionally he goes through a process of deliberation or reasoning, marshals evidence and principles, and draws conclusions. Nevertheless, if someone acts with an intention, he must have attitudes and beliefs from which, had he been aware of them and had the time, he could have reasoned that his action was desirable….  

Von Wright agrees: “In order to call a movement intentional it is not necessary that the subject should have formed an intention to perform it before actually performing it. Sometimes this [forming an intention] happens.” Let us suppose that I went to a party but there was no process of reasoning. Should we say that if I had been aware of my beliefs and had time before the party, I could have reasoned that I should go to the party? Perhaps, but this is not a satisfying answer to the question what s my reason to go the party was if I didn’t actually go through a process of reasoning. Davidson admits that an agent does not have to be aware of the reasons on which he acts and which can be used to explain his actions. But this being the case, what can ensure that the agent does not invent his reasons afterwards? My question is: what is the explanatory role of retrospective reasons on Davidson’s and von

---

1493 Davidson, 1978, 85.
1494 Von Wright, 1998, 142.
Wright’s accounts? That von Wright accepts retrospective reasons as genuinely explanatory becomes clear from the following:

Although an explanation in terms of reasons may point far beyond the context of the action in time, the reasons must yet, all of them, be present in the context. The agent need not be aware (“thinking”) of (all of) them when he proceeds to action but they must be present in the sense that he subsequently can say, if challenged, that he had them then. He did not invent them afterwards, nor had he completely forgotten about them. He would have been able to state them when proceeding to action had he, for whatever reason, reflected on why he was doing what he was doing.  

For behavior to count as intentional action, it is enough according to von Wright that I can give reasons for my action when challenged. These reasons can be such that I do not have to be aware of them when I begin to act, but they do not count as reasons if I invent them afterwards or if I had completely forgotten them. In other words, retrospective reasons which I can give when challenged count as explaining my action if I don’t invent them.

This is all very tempting, but here the problem that I am raising becomes clear. If my reasons for an action were reasons of which I was not aware before the action, how can we exclude the possibility that I did not invent them afterwards? I can say that I didn’t invent them, but I can be wrong. Human nature being what it is, we have a tendency to try to make sense of ourselves. This is precisely what happened in the experiment of Nisbett and Ross. The participants were not willing to admit that they had invented their reasons. If I give a retrospective reason and someone challenges it by saying “No, you didn’t have that reason, you invented it!” I will likely protest. I can sincerely declare that I had this reason before the action, although I wasn’t aware of it. Who is to say that I am not correct in my view? Well, here we have a reason to challenge my view on the grounds that if I was not aware of my reason, what actually is my reason to say that I had it before the action took place?

How would von Wright or Davidson describe the pantyhose situation? We should recall what von Wright says:

[the reasons] must be present in the sense that he subsequently can say, if challenged, that he had them then. He did not invent them afterwards, nor had he completely forgotten about them. He would have been able to state them when proceeding to action had he, for whatever reason, reflected on why he was doing what he was doing.  

Suppose that I am not aware of my reason to pick up pair D but I can state it when challenged. On the von Wrightian account this would count as a genuine reason if I had been able to state
it when proceeding to action. But what can ever guarantee that I would have been able to state it? Von Wright notes that I need not be aware of the reason but he also notes that, in order for my reasons to count as reasons, I should have not completely forgotten them. How can these views be reconciled? It is plausible that people firmly claim that they had a reason for what they did and they are “absolutely sure” that they did not invent it afterwards. This being the case, von Wright’s account cannot help to solve the question of whether what happened was an action of the agent. The same is true of Davidson’s position: “[…] if someone acts with an intention, he must have attitudes and beliefs from which, had he been aware of them and had the time, he could have reasoned that his action was desirable…. “1497 It is not required that I must be aware of my attitudes and beliefs if I act intentionally, only that had I been aware of them I could have seen the connection between the attitudes, beliefs and the action.

In a sense, Davidson and von Wright do worry about the questions of whether retrospective reasons can be efficacious and whether these reasons can be taken to refer to anything at all. It seems that von Wright is more aware of these problems than Davidson. He makes the distinction between existing and efficacious reasons and concludes: “Only of efficacious reasons do we say that the agent acted for those reasons or because of them.”1498 Only efficacious reasons belong to the explanation of person’s action in terms of reasons. This is an important distinction, but it does not separate reasons in the same way that the distinction between precursory and retrospective reasons does. The problem is that it seems possible that the retrospective reasons by reference to which we often explain can be non-existing and therefore non-efficacious. Davidson puts little emphasis on the view that we should be consciously aware of the reasons on which we act. As he notes: “It is clear that most of our actions are not preceded by any conscious reasoning or deliberation.”1499 This being the case, Davidson seems to accept the view that usually we explain our actions and the actions of others by referring to retrospective reasons. Von Wright has a similar view: “In many, perhaps most, cases the agent just has some reasons and then acts – and only in retrospect, if at all, does he reflect (“think) on them and makes it clear to himself or to others why he acted.”1500 This is what happened when I went to the party and gave a reason when

1497 Davidson, 1978, 85.
1499 Davidson, 1987a, 107.
1500 Von Wright, 1998, 142. This can however be compared to the following: "That an agent acted for a certain reason normally means that something was, for this agent, a reason for doing something and that he set himself (chose, proceeded, maybe upon deliberation) to do this thing for that reason." (Von Wright, 1998, 11) The problematic expressions are: "agent has a reason" and he "chose to do a thing for a certain reason." How should we understand the expressions "having a reason" and "choosing" if it is not required that the agent is aware of the reasons he has and if he does not consciously choose? Is choosing (and here I mean real choosing) without
asked. Perhaps I had several reasons and I possibly acted because of one or more of them, but I became aware of them, and I started to think about them, only when challenged. This is also what von Wright and Davidson could say about the pantyhose case; a person just had some reason and acted, and only in retrospect does he think of the reason and make clear to others why he acted. But this will not work if the “choosing” of pair D was really caused by its location.

Although von Wright claims that perhaps in most cases the agent just has some reasons and acts accordingly, he also thinks that normally an agent knows what he did and also why he did it; he knows what his reasons were. This being the case no explanation is ever required for most actions. Upon self-reflection an agent would straightforwardly know why he did a certain thing and nobody would have a reason to doubt this explanation. Although this is a common-sense view, it is naïve from a philosophical and psychological point of view, especially if explanations in terms of retrospective reasons are accepted as genuine explanations. The trustworthy self-reflective retrospective processes to which von Wright refers can be misleading. Von Wright is aware of this: “The self-reflective process can be called “rationalization”. Sometimes it consists in the agent just inventing, post hoc, reasons which in fact he did not have.” Von Wright’s exact definition of rationalization is: “The phenomenon called ‘rationalization’ occurs when I give reasons which were in fact not the reasons for which I acted.” Whereas in Davidson’s terminology “rationalization” is a form of explanation, one could interpret von Wright as claiming that rationalization is not a form of explanation. Von Wright’s view is not clear, though, because he notes that that the borderline between pre-existing reasons and the subsequent rationalization of an action is often blurred. In saying this, he seems to imply that the subsequent rationalization is a form of explanation in terms of retrospective reasons; yet he also claims that rationalization is a process in which I give reasons which were not the ones for which I acted, and in this latter sense rationalization is surely an incorrect explanation if it deserves the name “explanation” at all. Let us understand rationalization in this latter sense from now on. The problem that I have posed for von Wright and Davidson can be thus summarized: are explanations in terms of retrospective reasons always rationalizations? If the answer is yes, then retrospective reasons are never

deliberation possible? Isn’t the term “choice” defined as something which one makes because of certain reasons, and can you do this if you are not aware of the reasons?

1501 One should remember the huge amount of empirical evidence showing people’s defects in their self-understanding. Von Wright notes that the issues considered here create interesting problems for psychology but he ignores them as not being philosophical problems. In this context I am not certain how von Wright draws the philosophical / non-philosophical distinction and what he means by it.

1502 Von Wright, 1998, 142.

1503 Von Wright, 1998, 89.
genuinely explanatory. If the answer is no, how can we distinguish genuine explanations in terms of retrospective reasons from rationalizations?

Although I hesitate to conclude that explanations in terms of retrospective reasons are *always* pseudo-explanations because they do not give the reasons for which I actually acted, it seems difficult to see what could convince us that retrospective reasons really were reasons for my action. There seems to be a clear difference between cases when I am aware of my reasons before the action and *therefore* act, and cases where I, when challenged, give reasons of which I was not earlier aware, retrospectively. In the first case “the reason was there”; this is so to speak an ontic truth, and the main question is whether it was efficacious. In the second case we have to consider both whether the reason existed and whether it was efficacious. From the fact that something is a reason for action it does not follow that an agent has that reason for his action. From the facts that something is a reason for action and that the agent has that reason for action it does not follow that he does the action for that reason. Both von Wright and Davidson are of course aware of these facts, and their respective views in the philosophy of action are meant to solve the problem of how efficacious reasons could be pointed out. It seems to me clear that a reference to causality does not solve this problem. Von Wright, who sees no hope in the causal answer, argues that necessary criteria for a true reason-based explanation are twofold. First, the given reason must *be* (i.e. must count as) a reason. A physical event under its physical description would not count as a reason since having a reason requires understanding something. Second, the agent must have or must have had the reason on the occasion in question. My claim is that this approach cannot solve the question of whether retrospective reasons are efficacious. But is there any way to solve this problem? I think that we should applaud von Wright’s answer to the question of how to identify efficacious reasons:

The efficacious reasons are those in the light of which we explain the action. I maintain, in other words, that one cannot separate the question of the efficaciousness of the reasons from the act of understanding the action as having been performed for those reasons. *This means that the truth of the action explanation has no basis in facts other than the understanding itself of the action in the context of its reasons.*

In the spirit of what has been said about reasons and actions, this kind of view was to be expected. Mental concepts to which actions belong are so connected to each other that we cannot begin discussing the efficaciousness of reasons outside the context of the actions which they are taken to be reasons for. This observation of von Wright should be considered...
very seriously because it clarifies remarkably well what we mean when we say that the mental is a conceptual category. Mental phenomena exist only in a certain kind of context. We can make sense of the existence of mental phenomena only in a certain kind of context. What von Wright says in the above quote has the wider consequence that the efficaciousness of a reason is tied to the idea of a consensus about the question of which reason explains the action in the best way. Suppose that we have competing reasons for why I did something. Since equating a reason with a cause is out of the question, the only alternative is to see the explanatory – and therefore efficacious – reason as being the one which fits best the overall context of action. The efficacious reason is the one in terms of which the whole situation makes the most sense. This being the case, understanding becomes a central concept because there is no truth about the reasons for actions outside the consensus, outside our understanding of what the reasons for a certain action were. Von Wright concludes that in the end the truth of an action explanation consists in agreement between the participants who evaluate which reasons are to be taken explanatory ones in relation to a given action that needs explaining.

We cannot study the efficaciousness of a reason outside the context of action. This is something which also Davidson would accept. When we explain an action in terms of its reasons, we imply that those reasons were efficacious. The causal aspect of reasons is, so to speak, derivative from the understanding of actions and reasons at the mental level. Whereas we who want to defend the autonomy and irreducibility of the mental can easily agree to the view which emphasizes the connection between explanation and understanding, one could still ask the question about the status of retrospective reasons. If we understand an action in terms of them, should they immediately be counted as being efficacious? Does understanding an action in terms of retrospective reasons establish their status as efficacious reasons? Is it enough to make the retrospective reasons efficacious?

The proposal which I want to make and which is not easily accepted is the following. We should reject the idea that retrospective reasons can be really explanatory and thus consider the possibility that actions in terms of these kinds of reasons are not actions at all. This may be a disturbing result. According to this proposal, behavior would count as action only if an agent is aware of his reasons before the action and proceeds to the action because of those reasons. This means, perhaps quite counterintuitively, that my getting up from bed in the morning would not count as an action of mine unless I had made a conscious decision to get up. This may sound intuitively implausible but I see no harm in calling actions only those occasions of behavior which result from conscious reasoning. This would mean that we act only rarely but, as the neuroscientific and psychological evidence could be taken to suggest,
this may exactly be the case. Perhaps one reason why this view may seem to go so strongly against our intuitions concerning the frequency of our actions is the fact that many of our daily actions have become automatic. The mind is often planning, anticipating, predicting or remembering. It re-evaluates the past scenarios or creates models of the future. Minds seem to have this possibility to wander and not stay in the present partly because we so easily manage to function without thoughts. When problems arise the conscious mind comes to rescue but otherwise we go on quite well without it in our daily tasks. Automatized action, as we could call it, is an interesting phenomenon because conceptually it belongs somewhere between genuine action and mere reactions.

We may also call these automatic occurrences of behavior actions, but my claim is that they are not genuine actions and it would be better to describe them as reactions. However, this would be a misleading term as well because an automatized action is an event which had a reason of which the agent was aware when he learned how to act in certain circumstances. Consider for example the fact that I usually turn off the light in the living room when I leave the house. This has become a reflex-like operation. I claim that this behavior is not genuine action because there is no precursory reason of which I am aware before I act. Yet, if asked why I turned off the light I could easily give an answer: I want to save electricity. A child who is learning this fact, i.e. why the light should be turned off is, at least at some point, aware of why he turns of the light and is thus acting. It does not take long for the child to become automatized in this task and then he ceases to act and starts to react. This is what we do most of the time; react. These reactions or automatized actions are very much like actions at the superficial level precisely because they have their reasons in the past. Since we are often aware of these reasons when we see another person reacting, we describe the reactions as actions. Given this well-entrenched, deep habit it is difficult to accept the view that genuine actions should follow only from precursory reasons.

Most of the time we act automatically and it is worthwhile to ask whether many of our “actions” would be better understood as being closer to complex reflexes. Since we are reluctant, and rightly so, to call actions those occurrences of behavior for which no reason can be given, we should consider whether the same skeptical attitude should be taken towards behavior which is explained by reference to retrospective reasons. The status of these reasons as efficacious reasons is obscure. By claiming that behavior counts as action only if it can be explained in terms of precursory reasons, I am claiming that the capability to act is closely

---

1505 For a discussion, see Bargh, 1999.
tied to one’s conscious awareness of oneself. In their respective ways both Davidson and von Wright acknowledge its importance when they insist that our capacity to have reasons is a result of our development in a social setting among language-users. I think we could say that our becoming conscious is the result of the same development. Both von Wright and Davidson connect the capability to have reasons to the capability to understand and, as was argued in chapter one, the connection between understanding and conscious awareness is an intimate albeit not a clear one. At the same time it should be noted, as I showed in section 2.5.1, that von Wright’s and Davidson’s views about the nature of the mental do not explicate the role of consciousness and they do not hesitate to describe mental phenomena in terms of dispositions, which are ultimately physical mechanisms. Interpretationism does not require attitudes to be definite entities. Although this may not straightforwardly lead to mental antirealism, it does raise questions about the status of the mental. I suggest that we should explore further the view that it is not possible to act at all without the self-reflective capacity which having reasons for actions is. By stressing the importance of precursory reasons we are stressing the importance of the use of this capability in our talk of actions and their reasons.

4.2.2.1 The consequences of taking precursory reasons seriously

In the previous section it was claimed that an explanation referring to a precursory reason has priority with respect to an explanation which refers to non-mental causes. What kind of consequences would this kind of view have? I have claimed that the new problem of mental causation is the question of how to tell whether a mental explanation is true. The suggestion is that such an explanation is true insofar as it refers to the precursory reasons which the agent sincerely takes to be the causes of his action. The importance of this suggestion is that it allows us to see ourselves more transparently. If you contemplate your reasons and arrive at conclusions accordingly, this is the highest degree of freedom that you can exercise as a human being. The suggested approach encourages people to observe themselves more thoroughly; the better one knows one’s precursory reasons and contemplates them, hence the greater the chance that one’s actions really are caused by his beliefs and desires. This is not a claim which could be proved or disproved empirically; it is a suggestion about how to think about the relation between reasons and actions. From the third-person perspective there is always going to be a fatal defect in this model, because an interpreter must rely on the sincerity of the subject’s reports. But insofar as the suggestion leads to increased self-
monitoring and self-analysis, the importance of the epistemological approach for the acting individual can be notable.

In trying to think about the possible consequences of taking retrospective reasons seriously, I have greatly benefited from the views of psychologist Timothy Wilson, who has claimed that although our behavior is largely manipulated by adaptive unconscious responses, we may train this mechanism so that it will better respond in the ways in which we would like act. Wilson suggests: “By being careful observers of our own actions, we can learn a lot about ourselves… if we want to change some aspect of our adaptive unconscious, a good place to start is deliberately to begin acting like the person we want to be.”\textsuperscript{1506} He also notes: “The difference between self-revelation and self-fabrication is crucial from the point of view of gaining self-knowledge.”\textsuperscript{1507} I believe that an approach focusing on the role of precursory reasons could be a small step towards a model which could start solving these problems. This kind of suggestion may contain severe conceptual pitfalls, but it seems to me that the overall idea could nevertheless be worth exploring. We can suppose that our daily behavior is largely automatic and driven by unconscious processes; we have only rarely precursory reasons for our actions. But we also have an image of the kind of person we would like to be, and we have an ideal how we would like to act towards others. In most cases we merely act without contemplation. But if we focus to develop our capability for self-analysis, it may turn out to be the case that our capability for making well-contemplated decisions in the future will increase. Then we would not be at the mercy of our adaptive unconscious responses, or to the extent that it still guided our behavior, it would work more according to the way that we want to see ourselves – as individuals who make decisions and are responsible for their actions.

In the cases when we explain ourselves retrospectively, the doubt concerning the efficaciousness of reasons exists in principle. The conviction that a thought has caused an action rests on a causal inference that is always open to question. It could be argued that this doubt exists also in the case of precursory reasons and therefore the distinction between two different types of reasons cannot help in deciding whether the latter reason was really my effective reason. I admit that also precursory reasons are threatened by the possibility of incorrect causal inference, but I nevertheless suggest that the process of contemplation (which is essential in the forming of precursory reasons) may convince the acting individual that, as a result of contemplation, he now really knows his reasons. Explanations in terms of reasons are

\textsuperscript{1506} Wilson, 2002, 203.
\textsuperscript{1507} Wilson, 2002, 206.
generally explanations which try to capture the agent’s point of view; it is this point of view which the agent can explicate by focusing on his precursory reasons.

My suggestion that explanations in terms of precursory reasons should override other explanations could be accused of “subjectivism” and of overtrusting our ability for self-observation. These are justified complaints, but I suggest that here we should also consider the question what kind of image of humans we want to hold. When the reasons for our actions are concerned, do we want to hand over the final verdict to outsiders? It could be done if we decide that non-mental explanations override explanations in terms of precursory reasons. Then only the opinions of outsiders would count, because we do not have access to the causes of our actions. Von Wright concluded that there is something tragic about the cases where the opinion of the outsider overrides the self-understanding of a person. I agree with this; we should thoroughly consider what kind of position we are willing to defend.

From psychological research we know many cases the conclusions of which have similar consequences for human autonomy as the pantyhose experiment. For example, studies suggest that referees tend to unconsciously show a bias towards athletes with red outfits. If this is true, one conclusion would be that the referee favored a certain athlete, not because he thought he was the best competitor, but because the evaluator’s visual system was biased towards red objects. Likewise, it is shown that when people are in groups, an individual’s willingness to help a person in need decreases. When asked why help was not offered, people do not refer to the size of the group. Statistically this is true. But what is the relevance of these studies when we consider the behavior of an individual? An approach like Nisbett’s and Wilson’s focuses on the behavior of a person and replaces the asking for reasons with an attempt to come up with statistically supported generalizations which will reveal the causes behind the behavior. An approach like this sees reasons as causes and attempts to show that the real causes are something other than the reasons to which the individual refers. A consequence of this could be a new approach to human behavior; on this approach, the concept of reason disappears.

Is this the kind of approach to human behavior that we are willing to defend? The studies on “biased-referees” have already led to suggestions that red clothes should be removed from certain sports. A new specific view about the nature of reasons and causes may thus have concrete effects on our policies. I believe, however, that here something is going wrong. When we are asking a person for reasons, we are not interested in the unconscious

\[\text{Hagemann, Strauß, Leißing, 2008.}\]
physical processes which may be taking place. We are interested in the person’s thoughts, desires and emotions. When we try to find the reasons of a person we may be, and often are, interested to see the situation from the agent’s point of view. We want to see what it was in the action that was appealing to somebody. When we consider whether an action was reasonable we are considering the question of what is reasonable according to our understanding, but we are also considering the unique perspective of another human being. When, in the midst of everyday life, we are asking why somebody is at a party we are not requiring him to advance a hypothesis or a theory about his inner processes. We may be interested in finding out the attitudes of others because we are interested to understand why he has acquired further beliefs; to know the beliefs of another person is to that extent to understand the person. In these cases we are not concerned with the discovery of causes but with the attempt to understand the person’s behavior as an action of a rational animal. This point is again and again emphasized by Davidson and von Wright, and it is curious how categorically it is ignored by those who want to equate reasons with causes.

Those approaches, be it Nisbett’s and Wilson’s psychological approach, or Bickle’s new wave reductionism, which attempt to equate reasons with causes and replace ordinary explanations of ourselves with scientific explanations make the severe mistake of suspecting that a scientific explanation reflects a discovery about the true causes of our actions. Because mental explanations and non-mental explanations are different, the former cannot be replaced by the latter. Mental explanations from a third-person perspective are based on the interpretation of an agent, and this is a procedure which has no counterpart in the sciences, which try to explain behavior without the use of mental terms. There cannot be scientific discoveries concerning our true reasons, and an approach which tries to replace reasons with causes should be seen, not as a scientific discovery, but as a proposal about how to think of human action. The experiments of Nisbett and Wilson show nothing about the reasons of the participants and it cannot turn out that explanations in terms of causes can show that reason-based explanations are false or can be replaced, as for example the Churchlands or Bickle claim. Whether we are willing to accept a new way of thinking about human behavior is a question the answer for which depends on what kind of value we want to put on the reasons that other human beings offer for their actions. The kind of value we will choose to put on these reasons will also show what our relationship to other people is and what our attitude towards another thinking being is.

If we wish to know why a referee judged the red-suited athlete to be the winner, we are not looking for the answer that his visual system shows a bias towards red objects. In a
society where the importance of science has been overemphasized at the cost of understanding causal concepts, a tendency to explain with only one type of cause (a tendency against which already Wittgenstein warned) has become a trend. If this trend continues and escalates, we may arrive at a point where the causal hypotheses start to concretely compete with rational explanations. Then the choices we have to make are real, practical and very pressing. By suggesting that precursory reasons should have authority over causal ones, I try to point us in a direction where such conflict would not occur. Obviously, those who would be willing to misuse the causal hypotheses which are backed up by empirical evidence are not moved by the sincere reports of an agent. As von Wright pointed out: “It is characteristic that those who misuse their authority when they disqualify the testimonies of the agents often do this in the name of ‘higher’ truth, perhaps sanctioned by ‘science,’ which the recalcitrant agent is been forced to accept.”\textsuperscript{1509} It seems to me that if the causal explanations start to triumph over rational ones, this will be a time when scientific discoveries result in the destruction of human beings. It is a start of very unfortunate era.

4.3 Conclusions of chapter four

The discussion of this chapter is based on the conviction that non-reductive physicalism cannot offer a reductive solution to the problem of mental causation. This conviction has remained largely without argument, but I fail to see how a position which strongly emphasizes in-principle irreducibility could come up with or accept an explanation that explained the causal efficacy of the mental in neural, or more broadly in physical, terms. This being said, at least on Davidson’s non-reductive account causal efficacy is easily explained. In sections 4.1, 4.1.1 and 4.1.2, I described two alternatives by which Davidson might secure the causal powers of mental phenomena. On the one hand, mental events are identical with physical events and physical events have causal powers. On Davidson’s account, identity grants causal powers to mental events; physicalistic monism secures the causal efficacy of mental. On the other hand, mental events are events and events have causal powers. Both answers are unsatisfactory from the critics’ point of view. They would reply that both answers leave unexplained the question of how exactly mental phenomena cause – for example, how it is the case that the content of a belief can have causal efficacy. It seems to me that this question remains, and must remain, without an answer if a non-reductive position is accepted.

\textsuperscript{1509} Von Wright, 1985, 27
Von Wright remained outside the contemporary mental causation debate. I argued in section 4.2.1.1 that his most interesting contribution to the discussion was the consideration about the status of epiphenomenalism. Although von Wright’s final word was that “of course” reasons move agents to act and that mental explanations enjoy autonomy over neural ones, I believe that his discussions of epiphenomenalism may provide an interesting starting point for those who see a form of epiphenomenalism as a real possibility. I suggested that a non-reductive position cannot avoid the possibility of conditional epiphenomenalism, and that this possibility is something which should be accepted as part of our situation. A suggestion about how apparent mental causation could occur was given in section 4.2.1.2.

A first sketch towards a new way to see the problem of mental causation was offered in section 4.2.2. This “solution” of the new problem of mental causation starts from an epistemological of deflationary perspective, thereby rejecting the overtly ontological approach to the problem. By stressing the importance of precursory reasons the model takes into account the view that a satisfactory account of mental causation must grant a causal role to states of the brain. My claim is that Davidson’s and von Wright’s positions do not explain this role in a satisfying way because these positions do not require that an agent be aware of the reasons which explain his actions. This being said, by focusing on the ways in which mental explanations work, the approach that focuses on precursory reasons privileges mental facts over neural ones. Section 4.2.2.1 contained a brief discussion of the possible consequences of the view that precursory reasons are required for genuine actions. Two consequences should be emphasized. On the one hand, by consciously focusing on the reasons one has for one’s actions, one may influence one’s adaptive unconscious, thereby getting a better control over one’s automatized behavior. Although complete escape from self-determination is an idle dream, my claim is that the better one is aware of one’s reasons the more freedom one enjoys. On the other hand, an approach which emphasizes the epistemic practice of explaining an agent’s actions in terms of his precursory reasons is a suggestion about how we should see the nature of human being; as an autonomous being and not a machine, the behavior of which can be explained in purely physical terms.
Appendix: Against Material Mind

According to ancient wisdom it is better to travel well than to arrive. Sometimes the process of reaching a conclusion may be more important and edifying than the conclusion itself. The writing process of this thesis led me to consider those broader problems of physicalism that should concern us in the future. After the thesis was completed I realized that the focus of this work should have been on different questions, but by then it was too late to change anything. In this appendix I very briefly turn my focus to the kind of problems which, or so I believe, should be brought into the discussion in the contemporary philosophy of mind.

My original motivation for writing this thesis was to clarify Davidson’s and von Wright’s positions, to understand their claims better, and to consider whether these claims are relevant in the context of modern naturalism and its physicalistic ontology. Whereas the writing process made Davidson’s and von Wright’s positions clearer, a conviction that these philosophers were not really addressing the deepest problems of physicalism started to emerge. My initial impression was that von Wright and Davidson could resist the challenge of naturalism because of their understanding of the nature and purpose of philosophy. In chapter one, I showed that their conception of philosophy is influenced by Wittgenstein’s view that philosophy is not one of the sciences. However, when we consider Davidson’s and von Wright’s arguments in the philosophy of mind, I think we have to conclude that in many cases they try to compete with empirical views and are not faithful to their conception of philosophy.

Wittgenstein once asked: what is the use of studying philosophy, if it does not improve your thinking about the questions of everyday life. He thought that it was very difficult to think really honestly about one’s own life and other people’s lives. These considerations are more difficult and more important than the considerations about abstruse philosophical questions. Wittgenstein concluded that the trouble was that thinking about the questions of everyday life was: “...not thrilling, but often downright nasty. And when it’s nasty, then it’s most important.”1510 I think it is these nasty questions on which we should increasingly focus as philosophers. In times, like ours, when the belief in scientific progress is well accepted in society and has mesmerized many philosophers as well, we should critically evaluate the view that only science will carry us into an earthly paradise. It may be intellectually amusing, perhaps even satisfying, to challenge other philosophers’

1510 As quoted by in Malcolm, 1984, 93-94.
interpretations or show that there are inconsistencies in the views of, for example, Davidson and von Wright. In a doctoral thesis it is necessary to show that new “scientific knowledge” has been produced. All the “technical work” in this thesis, such as the analysis of how von Wright’s supervenience should be understood or what Davidson exactly means with his argument for the anomalism of the mental, is very entertaining. But are we, in focusing upon these technical questions, turning philosophy into the playground of professional philosophers where the discussion and argumentation goes around in circles? Are we turning philosophy into a self-sustaining enterprise with no practical import? Has philosophy become a specialized and compartmentalized discipline where there is no room for considerations of how to live a decent life or what kind of idea of human beings we want to hold? Contributing to the technical discussion that is carried on by professional philosophers and convincing them that one’s research is “vitally important” for the contemporary discussion is the best way to get funding for one’s research. But is this kind of philosophy important, i.e. does it matter? Should we, instead of letting the philosopher continue these inbred discussions, demand that (s)he should work also on unhealthy habits of thought, unchallenged assumptions, and values that are taken for granted? Should philosophers reach beyond their narrow discipline and address those problems which influence the majority of humans? I believe, without attempting to prove it in this work, that the answer to these questions is affirmative. This belief, which is the result of this work, casts serious doubt over the previous five hundred pages of description and analysis.

Instead of discussing the technical details of why physicalism is a problematic position, we should consider why it may be a nasty view – and thence focus on the possible consequences to which this ideology may lead. Yes, I describe physicalism as an ideology which reflects the deeper motivations of its practitioners. Trying to show the connection between a philosopher’s philosophical views and their views, for example, on politics is a formidable task, and one which I shall not attempt here. I claim, however, that the most serious threat of a material mind is that many of its prophets also have views about the direction in which society should be heading, and the material mind is an auxiliary which helps them to reach this destination. If we want to object to this destination, we have to fiercely object to the material mind.

In this appendix I am not trying so much to argue for a view as to raise thoughts and concerns. Perhaps it is a mere coincidence that a materialistic view of human nature is in many cases related to certain kinds of societal and political ideas. Without insisting that these views are related, I merely give a few examples of how broader questions are guiding any
philosophy of mind which has turned naturalistic. The purpose of these examples is to challenge the reader to consider whether a physicalistic philosophy of mind could be unhealthy in some ways, and whether the ideology behind it should be challenged. Certain forms of reductionism in the philosophy of mind go together with the dehumanizing trend of modern society. The desire for mental-physical reductions can be seen as an expression of how vocabularies should develop. Reductions pave the way for conceptual changes. These changes will eventually lead to more substantial changes. When forms of speech become part of the ordinary way of talking, they shape our world. When another human is a brain instead of a person to us, the consequences go far beyond technical philosophical disputes. A scientific conception of human beings that is based on the results of brain research may create concrete problems that decrease the sphere of human freedom and autonomy. I believe that the most serious challenge for physicalism is thus a kind of ethical concern, relating to how a physicalistic ideology shapes our ways of thinking and thereby human life.

Neurophilosophy is an updated version of naturalism, one which challenges also conceptual analysis (philosophy as it is traditionally understood). Conceptual analysis has “stumbled to its knees”, as Patricia Churchland so vividly claims. Her view is that productive philosophers of mind/brain are steeped in the relevant empirical sciences. I would claim that since Western society has a very positive attitude towards science and progress, it is not surprising that non-empirical philosophers of mind are currently a minority. In the prevailing atmosphere of optimism it is unfortunately often forgotten that, when the term is used in a value-free manner, progress may lead to disaster. Scientific philosophers, sometimes showing open contempt for traditionally oriented philosophers, claim that the results of their research have only positive consequences and that an agenda which reveals objective facts cannot be harmful. “Truth” must be allowed to take its course. It is easy to see, however, that neurophilosophy is a project with potentially serious threats, and it could be argued that the seeds of these threats are already visible. Science may turn out to be a threat, not only to philosophy but more generally to human life. In his recent book Neurophilosophy at Work Paul Churchland suggests that cognitive neurobiology and computational neuroscience can solve many of the perplexities which have troubled philosophers. Churchland suggests, for example, that cognitive neurobiology could have an impact on moral issues, which have been the concern of philosophers, and that it could serve as a tool in moral correction:

1511 Churchland, 2008.
There remain... the genuine abnormals, for whom moral correction is first a matter of trying to repair or compensate for some structural or physiological defect(s) in brain function. Even if these people are hopeless, it will serve a social policy to identify them reliably, if only to keep them permanently incarcerated or otherwise out of the social mainstream... Where the deficit is biochemical in nature - giving rise to chronically inappropriate emotional profiles, for example - neuropharmacological intervention, in the now familiar form of chronic subdural implants, perhaps, will return some victims to something like a normal neural economy and a normal emotional profile... these individuals will then also be candidates for the resocialization techniques imagined earlier for disadvantaged normals.  

I hope that the reader shares my uneasiness with this view. Why is the view dangerous? When a neurophilosopher refers to a future brain science and in the same context refers to hopeless people who have chronically inappropriate emotional profiles who should be taken care of with neuropharmacological intervention, he is putting forward an agenda. As part of this agenda, Churchland has speculated how future generations may see the world in such a different way that they:

[...] sit on the beach and listen to the aperiodic atmospheric compression waves produced as the coherent energy of the ocean waves is audibly redistributed in the chaotic turbulence of the shallows. They do not observe the western sky redden as the Sun sets. They observe the wavelength distribution of incoming solar radiation shift towards the longer wavelengths (about 0.7 x 10^-6 m)...  

These generations will communicate with “Übersetzen” or through direct brain-to-brain links. Churchland suggests that when science reveals truth about reality, children can be trained accordingly so that they will enjoy the pleasure of seeing reality in a true perspective. The claim that a specific group of people knows what is good for others is familiar from the course of history. Should we worry that the brave new world of neurophilosophy, where inappropriate emotional profiles are fixed, may turn out to be a disquieting place? Once the view that neuroscience describes the one true human nature is accepted, it becomes easy to think of each other as mere brains instead of equal persons who deep down share a system of beliefs and values. When neuroscience makes the demarcation between normal and genuinely abnormal, how many hopeless people will we find?

Churchland believes that the development of neuroscience will: “[...] reconfigure our legal practices, our correctional practices, our educational practices, and perhaps even our recreational and romantic practices.” But it is not obvious that a technologically more powerful theory of human nature is necessarily an improvement. Whereas Churchland admits that the benefits of neurophilosophy are not certain, he does not consider the possibility that

\[1512\] Churchland, 1998e, 52.  
\[1513\] Churchland, 1979, 29.  
\[1514\] Churchland, 2000, 74.
the program of neurophilosophy could worsen our situation. We can control the outcomes of science because, according to Churchland, individual people of our age are at least “little superior” both in scientific and moral knowledge when compared to the people of the previous centuries.\textsuperscript{1515} We Westerns know that earth is not flat and we do not burn people as witches. Our conception of what is socially acceptable has changed for the better. The possibilities of what we can become have increased enormously. A person can be a real estate agent, a securities investigator or a congressional lobbyist. Churchland thinks that these activities are new contributions to the well-being of humankind and the increase of scientific and moral knowledge has made this progress possible. The existence of corporations, stock markets, and banks are signs of progress suggesting increases in the quality of life. Thanks to science, now hand-held GPS devices and credit cards are possible.

This is what Churchland counts as progress, progress for the better. If uncontrolled growth is progress, then – I guess – humankind is making progress. In the current system of global capitalism, the desire to “improve” things is always present. In this kind of competitive environment different providers, be they philosophers or salesmen, see their product as the product which makes life better. It is telling that in Churchland’s view: “[…] we’re coming to an age where cognitive prosthetics, whether they’re, you know, just sticking some extra RAM into your head somewhere, or more realistically, interacting with a machine. We can enhance human cognition….”\textsuperscript{1516} It is indicative of the current ideology that cognitive prosthetics are seen as an enhancement of human cognition. Would extra RAM in our heads improve our lives? Perhaps it would allow us to perform ordinary tasks more effectively than we currently do. But whose interests would this serve? Is it an objective worth pursuing? It is in the modern capitalistic system, where the value of a human being is measured by how well (s)he “performs” or achieves the goals set by the society.

Philosophers do not work in an intellectual vacuum outside society. A philosopher’s world view’s relation to his philosophy is a complex issue, but I think a question worth considering is whether there is a connection between Churchland’s philosophy of mind, his conviction that scientific progress should continue, and his view that America is a great

\textsuperscript{1515} Patricia Churchland (2008) notes: “Moral attitudes can change when the benefits of a technology are clear and demonstrable.” This is true, but the change is not necessarily for the better. Of course, from the viewpoint of an American scientist / philosopher certain customs, beliefs, habits, values or moral characters may look such that they are in need of revision. But this person lives in a society where certain values, beliefs, and moral characteristics are taken to be exemplary, and where the belief in scientific and technological progress and solutions is high.

As a curiosity we can note that Patricia Churchland has written a chapter for *Human Dignity and Bioethics: Essays Commissioned by the President’s Council on Bioethics*. The council advises the US president on ethical issues which are related to advances in biomedical science and technology. Churchland’s advice for the council is: “[…] far from being undermined by neuroscience’s insights into human behavior and its causes, moral responsibility is actually put on a firmer and more realistic basis, the more we understand about the neurological substratum of our moral life.” In his book *The Ethical Brain* Michael Gazzaniga, one of the pioneers in cognitive neuroscience, sees neuroethics as: “[…] an effort to come up with a brain-based philosophy of life.” He thinks that modern science should ultimately replace “non-rational belief systems” and describe a “natural order” in which we all can believe. Gazzaniga hopes that neuroscience will reveal the ethics built into the human brain so that we may begin to live more fully by them.

Is it a coincidence that Gazzaniga is also part of the US President’s Council on Bioethics? He is not worried about the negative consequences of scientific progress because: “[…] society as a whole seems always to return to the reasonable use of new knowledge.” When considering whether cloning-technology could lead to drastic negative results Gazzaniga concludes that it could not: “[…] because we are a moral society that will not allow such extremes. While they have occurred throughout history, we have gotten rid of them—whether they be extreme dictators, extreme fashions, or extreme drugs.” If we consider the history of the 20th century this view sounds absurd and almost like an insult against those who suffer in the world today. Gazzaniga’s view cannot be based on anything but faith that moral progress goes together with technological progress. Of this we have not one shred of evidence. Gazzaniga’s view about scientific progress is not based on science but on his interpretation of human history and on the morality of American people.

Is it a co-incidence that Bickle – who demands ruthless reductionism, does not give a hoot about the philosopher’s intuitions and does not bother to read Kripke – is “a self-described political conservative” who opposes the “poverty and silly demands most Ph.D. graduate programs impose on students” and argues that the solution is to point “some conservative undergraduate students to Ph.D. programs and exposing them to the joys of

---

1517 The same question should of course be asked in the case of other philosophers; I use the Churchlands here only as an example.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
1519 Churchland, 2008a, 117.
It is surprising that a conservative neurophilosopher defends reductionism instead of methodological pluralism? Is it a coincidence that William Casebeer a philosopher, cognitive scientist and an intelligence officer in the United States Air Force, who is one of the writers of the report “Neuroethics and National Security”, has published with Patricia Churchland and received “the campus-wide outstanding thesis award” for his dissertation completed at University of California at San Diego, where both Paul and Patricia Churchland are professors?

In sum, is it a coincidence that there seem to be many relations between neurophilosophy and certain kind of politics, the worst example being academic involvement in military research? Why is it that a neurophilosopher is sitting in the president’s council instead of non-reductive physicalist, panpsychist, emergentist, or epiphenomenalists? There are of course clear reasons why a neurophilosopher is heard over epiphenomenalists when it comes to political and societal questions, but this also shows that in modern society there is a need for a certain kind of philosophy of mind. I am not claiming, obviously, that every philosopher of mind with a reductionist view is a political conservative. I am not seeing a conspiracy of conservatives who try to infiltrate the philosophy of mind. I am merely giving a few examples of how positions in the philosophy of mind go, if not hand in hand with politics, at least quite well together with specific ideologies. Gazzaniga thinks that America is a moral society and Paul Churchland notes: “Thanks first to our founding constitution, America has been free of the sorts of spiritual and intellectual oppressions often found elsewhere.” The spiritual-cum-political tyranny displayed in “fundamentalist Muslim countries such as Afghanistan and Iran” is unthinkable in America, claims Churchland. He also thinks that America is one of the best examples of religious diversity in human history and believes that is one of the reasons why America is a great nation. Without taking a stance on these opinions, I believe we can justifiably ask how they influence the philosophical positions of their holders. Perhaps Churchland believes that a terrible misuse of neuroscience is unthinkable in the United States. But given the moral example that America has shown, this belief is certainly an illusion.

It is not obvious that scientific progress is something that will benefit mankind. The agenda of neurophilosophy is, however, based on the conviction that an increase of scientific

---

1524 Bickle’s brief views about conservatism are from: http://www.danielpipes.org/comments/21631.
1526 For a discussion of the relations between brain research, academics, and the military see Moreno, 2006.
1527 Churchland, 2001, 76.
1528 ibid.
knowledge is good in itself. Paul Churchland writes that because sciences which study the brain are: “[…] pregnant with promise and their effects on social practice are already being felt”\footnote{Churchland, 2000, 74.} it is therefore “[…] now hardly the time to become faint of heart or feeble of vision.”\footnote{Ibid.} This is the talk of a demagogue; we must move forward and there is no time to hesitate. Paul and Patricia Churchland raise their criticisms especially against religions, the hinderers of progress, which try to impose their moral truth on everybody by insisting that: “[…] I have to dress in a certain way, or I have to have my feet bound or I have to wear a bag over my head or I have to have such and so many babies or I can’t have any babies or blah and blah and blah….”\footnote{From the transcript of an interview, retrieved from \url{http://thesciencenetwork.org/programs/the-sciencestudio/from-the-engine-of-reason-to-the-seat-of-the-soul-a-brain-wise-conversation}} I find this pejorative talk especially disturbing since it is coming from esteemed academics. I find it worrying that neurophilosophers who downgrade the beliefs of people living, for example, in “fundamentalist Muslim countries” are eager to talk about genuinely abnormal people who should be fixed with resocialization techniques or neuropharmacological interventions. The Churchlands object to the idea that metaphysics could give a special insight into moral truths. But scientific and political enterprises are not really that different from religious fanaticism. The science of today tells how people should live and imposes its metaphysics into the moral sphere. If neuroscientists and neurophilosophers are sitting in the President’s Council on Bioethics of the world’s only superpower, it is reasonable to think that their metaphysical beliefs partly dictate policies which have effects on other people’s lives.

I believe that Patricia Churchland’s observation applies also to the enthusiasm of neurophilosophers: “Even thoughtful, experienced, balanced people may be ignorant of certain facts or may themselves be blinded by certain hopes and passions.”\footnote{Churchland, 2008.} The best way to keep the neuroscientific enthusiasm alive is to emphasize its benefits and ignore its negative aspects. This strategy has been used many times before, when it has been important to sell a new scientific development or ideology to a wider audience. It is commonplace in science. Research in physics was not advertised by emphasizing the fact that it perhaps allowed the development of weapons which are now a threat to our existence. Gene research is not marketed saying that it may accidentally create new diseases.\footnote{Such cases have already occurred, and in retrospect it is now known that the scientists responsible were hesitating on whether the results should be published or not. In any case the damage was already done.} I haven’t seen neurophilosophers discussing the possible abuse of their brain-centered conception of man.
The possible serious consequences for our self-conception and for human freedom will never be the consequences in terms of which brain research will be sold to the public.

I believe that if we want to consider why a reductionist view about the mind is really a threat and why a material mind should be objected to, we have to consider the harmful consequences of reductionism, the ideology on which it is based and the motivation which drives its proponents. Likewise, if we want to understand why Davidson and von Wright defend a non-reductionist and in some sense unscientific position in the philosophy of mind, we must understand their motivation for defending this kind of view. My claim is that the motivation has nothing to do with the technical philosophical arguments involved, but is closely related to von Wright’s and Davidson’s views about the nature of a humane society and about the idea of human beings that they are willing to defend. This motivation is not very different from Wittgenstein’s skeptical attitude towards scientism and the dehumanizing trend of modern society. I will point out only the sources of motivation of these three philosophers, suggesting that this is a topic for further future consideration and research.

Wittgenstein’s non-philosophical background strongly influenced his philosophical views. The relation between Wittgenstein’s Spenglerian view about culture and his philosophical views is much discussed. Wittgenstein’s influence on von Wright is also well recognized. The relation between von Wright’s philosophy of mind and his Wittgensteinian pessimism has not, however, been studied well enough. Wittgenstein himself feared that the more we know scientifically about human psychology, the less secure could our understanding of others and ourselves be. In 1945 Wittgenstein wrote to Malcolm: “Only extraordinary scientific achievements have a way, these days of being used for the destruction of human beings. (I mean their bodies, or their souls, or their intelligence). So hold on to your brains.” 1535 This was a result of Wittgenstein’s more general worry that possibly:

…the age of science and technology is the beginning of the end of humanity; that the idea of great progress is a delusion, along with the idea that the truth will ultimately be known; that there is nothing good or desirable about scientific knowledge and that mankind, in seeking it, is falling into a trap. It is by no means obvious that this is not how things are. 1536

This pessimism about the idea of scientific progress was likely fuelled by Wittgenstein’s experiences during the first decades of the 20th century. He was not optimistic about the prospects of science: “Nothing is more conservative than science. Science lays down railway

1536 Wittgenstein, 1980a, 56e.
tracks. And for scientists it is important that their work should move along those tracks."\textsuperscript{1537} It is characteristic of science that it elbows out other intellectual methods, ignoring the purposes of these methods. Wittgenstein’s words sound prophetic once we consider how scientific philosophy has become a dethroner of traditional philosophy.

How should we understand the claim that extraordinary scientific achievements could be used to destroy the soul of the human being and that we should therefore hold on to our brains? I think two readings can be given. First, the suggestion that we should hold on to our brains could be taken to be merely a plea for healthy skepticism against the hegemony of science. Wittgenstein thought that philosophers are often tempted to imitate the methods of science and this desire has foul consequences. So, perhaps Wittgenstein was merely suggesting that thoughts should be kept honest and critical. He made the comment when discussing psychoanalysis; although Wittgenstein admired Freud, he also claimed that false conclusions could be drawn if psychoanalysis was understood as a science which describes how things really are. The hypotheses of psychoanalysis should not be seen as scientific claims but rather as concepts offering a new perspective on human nature. In section 4.2.2.1, I made a similar remark about the “causal view” that equates reasons which causes and, like in Nisbett’s and Wilson’s experiment, tries to replace incorrect rational explanations with explanations referring to causes. The causal explanations are based on statistically supported generalizations. The suggestion that these causes should replace an agent’s reasons is not a scientific discovery; it is rather a proposal about how to think of the nature of human beings.

I think a more interesting reading of the “hold on to your brains” claim is that Wittgenstein wanted to warn against neuroscientific infiltration of the brain; this warning is more relevant now than it was in Wittgenstein’s times. What is the reason for this warning? Wittgenstein’s thoughts on this question are hard to follow. I think one important suggestion is that a science which focuses on the brain and perhaps attempts to replace ordinary mental concepts would impoverish our ways of reacting to each other. Wittgenstein wrote: “Psychological concepts are related to those of the exact sciences as the concepts of science and medicine are to those of old women who spend their time nursing the sick.”\textsuperscript{1538} The concepts of old women are important because they serve different purposes than the concepts of science and medicine; when we encounter a sick person we should not care just about the disease but should also pay attention to the human being who has the disease. Sometimes doctors confront a disease without treating the patient; in the age of medicalization people are

\textsuperscript{1537} As quoted by Rheese, 1984, 202.

\textsuperscript{1538} Wittgenstein, 1980, 62.
offered simple solutions in terms of drugs. Perhaps doctors who cure but do not care do not have time or skills to use the concepts of old women. Medicine without comfort, which focuses on prevention or cure, is the medicine of our times. Wittgenstein thought that we should address other human beings in a different manner. He wrote to his friend Drury, who was practicing medicine: “Look at your patients more closely as human beings in trouble and enjoy more the opportunity you have to say ‘good night’ to so many people. This alone is a gift from heaven which many people would envy you. And this sort of thing ought to heal your frayed soul.” If a human being is for us merely an object for scientific study and if we do not study the faces of our fellow men closely enough we, according to Wittgenstein, do not live in the world in which we are. If we would look closely enough into the faces of others we would no longer think that they may turn out to be beings without beliefs and desires, and we would no longer pursue our reductive agendas; we would accept that certain facts are visible – like our lives – and would no longer crave explanations.

If psychological concepts relate to the concepts of the exact sciences in the same way that the concepts of old women do to the concepts of medicine, then it could be argued that psychological concepts serve essentially different purposes than the concepts of the exact sciences. It is these concepts that are essential in keeping the: “[...] faith …what my heart, my soul needs…”; it is psychological concepts that are for the “[...] soul, with its passions.” This sounds eloquent. But an aspect of our being escapes rationalization and eludes scientific exploration. In the current culture it is too easy, too convenient, to take a certain picture of what it is to be human for granted; according to this picture we are just very complicated machines. But the way we understand our own passions, faith or other psychological phenomena, cannot be corrected by claims which refer to neural facts. Is it naïve to hold on to this mystical dimension? Wittgenstein wrote: “In former times people went into monasteries. Were they stupid or insensitive people? - Well, if people like that found they needed to take such measures in order to be able to go on living, the problem cannot be an easy one!” To dismiss those aspects of human life which elude scientific solutions is a form of dehumanization; these aspects are essential part of human life and the psychological concepts which are needed for describing them stand in the middle of our life. “The human being is the

1539 Quoted in Rheese (ed.), 95-96.  
1540 Wittgenstein, 1980a, 33.  
1541 Wittgenstein, 1980a, 49e.
best picture of the human soul”, as Wittgenstein suggested. What happens to the human soul if we turn our gaze inwards towards the brain?

Material mind banishes the human soul. What kind of being is a soulless being? Wittgenstein thought that if we imagine a person without a soul, then we really have to imagine a body which acts like an automaton and not like normal human bodies, the essential feature of which is their unpredictability. The more we stress the importance of unconscious mechanisms in the explanations of actions, the more we see each others as machines and the less emphasis we put on the view that human beings are beings with souls – that is, beings that should be treated in a certain way. In Zettel Wittgenstein imagines a tribe that we would like to enslave. The justification for this would have been that: “The government and scientists give it out that the people of this tribe have no souls; so they can be used for any arbitrary purpose.” If anyone would claim that there is something mental going on inside these beings this would be laughed at “like stupid superstition.” It seems to me that something like this is our current inhumane attitude towards non-human animals; they are beings without souls. Perhaps in the future an attitude somewhat similar to this will be our attitude also towards other humans. Eliminative materialists ridicule the idea that we act on our beliefs and desires. They claim that our basic conception of human cognition may be a myth – stupid superstition – which has been moderately useful in the past, but which is false at the core. Neuroscientific evidence is ready to bring down this myth. It may also allow us to detect the people who are hopeless, beyond neuroscientific repair, and who should therefore be identified and kept permanently incarcerated. Are these people modern versions of Wittgenstein’s slaves?

One further dehumanizing consequence of brain research which worried Wittgenstein was that we would come to explain human action solely in terms of causes. A culture which emphasizes the usefulness of scientific knowledge is in danger of becoming a culture where there is a tendency to explain with only one kind of cause – that which is familiar from the natural sciences. It seems to me that Wittgenstein goes against this tendency when he writes: “No supposition seems to me more natural than that there is no process in the brain correlated with associating or with thinking; so that it would be impossible to read off thought-processes from brain-processes.” He continues: “It is thus perfectly possible that certain psychological phenomena cannot be investigated physiologically, because physiologically

\[1542\] ibid.
\[1543\] Wittgenstein, 1967, §528.
\[1544\] Wittgenstein, 1980b, §97.
nothing corresponds to them.”

This resembles the claim that token-identity is confusion; nothing physiological corresponds to psychological phenomena. Davidson, for example, thought that: “[…] human thoughts, correspond to so little of the material world that no self-contained science can be based exclusively on them.”

Wittgenstein’s claim that there may be a “[…] psychological regularity to which no physiological regularity corresponds” could be likened in spirit to Davidson’s claim against the possibility of psycho-physical laws. But Wittgenstein’s claim is more daring because he is suggesting that there does not need to be a physical cause mediating in the nervous system between psychological phenomena. As I have shown, von Wright tentatively accepted this view.

Wittgenstein though that: “Nothing is more important in explanations of thought and brain processes than throwing away all the prejudices about causality. This seems to me by far the most important step.” I believe we can conclude that, with some reservations, something similar could have been said also by von Wright and Davidson. Von Wright’s references to epiphenomenalism and his claim that consciousness stays outside the chain of causally related events suggest that in his view there is an essential difference between the causal relations in nature and the ‘causal’ relations involving mental phenomena. Davidson, on the other hand, suggested: “The fact that these relations [between the mental and the physical] are so deeply puzzling suggests some sort of misfit between the concept of causality and the concept of a thought.”

The claim that there is a misfit between the concept of causality and thought is not as radical as the claim that all prejudices about causality must be rejected when thinking about thoughts and brain processes. We are tempted to insist that mental causation must have a physiological explanation. But, as Wittgenstein asks: “Why don’t we just leave explaining alone? – But you would never talk like that, if you were examining the behaviour of a machine! – Well, who says that a living creature, an animal body, is a machine in this sense?”

My claim is that Davidson, like Wittgenstein, is worried about the possible dehumanizing consequences of reductive views. This worry is what ultimately motivates his non-reductive position. One of the most important and interesting interpretations of Davidson that I have come across is Bill Martin’s view that AM is not only a view in the philosophy of

---

1547 Davidson, 1993f, 312.
1550 Davidson, 1993f, 298.
mind but also an “ethical-political principle”.\textsuperscript{1552} I believe this is an accurate way to see AM in a wider setting.\textsuperscript{1553} It explains why in Davidson’s view the irreducibility of the mental is principled \textit{in the sense} that we cannot really \textit{allow} the reduction or elimination of mental concepts. In chapter three, I discussed how Davidson claims that a nomological slack between the mental and the physical is essential \textit{as long as} we want to hold a certain view about ourselves. By describing our behavior in terms of mental concepts, we remove it from the direct reach of any physical law. Davidson is not denying that increased knowledge about the brain would not better explain why we think, reason and act as we do. He did research in experimental psychology before coming to accept the view that psychology should be likened to philosophy. He was thus aware of psychological research and suggested that brain research can make relevant contributions to the problems of mind. There are thus reasons to think that the principled rationale for defending the irreducibility of the mental \textit{is something else} than the belief that mental phenomena are mysterious and beyond the reach of science. The reason to defend the irreducibility is deeper than most commentators are willing to see.

I think we should conclude that there may be forms of reductionism which can be compatible with Davidson’s views about the irreducibility of the mental. In a sense, the irreducibility is more relaxed than most critics have thought. As I concluded in section 3.1, the thesis of the anomalism of the mental suggests only that mental–physical laws cannot be as strict as the strictest laws of physics. But Davidson acknowledges that there are pragmatic reasons why the status of our mental vocabulary should be secured. He notes that the common-sense perspective of folk-psychology is a “bit freestyle”, but indispensable.\textsuperscript{1554} I think that this view can be likened to Wittgenstein’s suggestion that psychological concepts resemble the concepts of old women who take care of the sick. Davidson notes: “We have such a keen interest in the reasons for action and other psychological phenomena that we are willing to settle for explanations that cannot be made to fit perfectly with the laws of physics.”\textsuperscript{1555} Davidson admits that he can imagine a science concerned with a human behavior, one that would be purged of mental concepts, but notes that he cannot see what its interest would be; what would this science tell us about human action? This is Davidson’s pressing question.

The pragmatic considerations do not yet explain why AM should be seen as an ethical-political principle. It becomes clear only when we consider Davidson’s claim that the

\textsuperscript{1552} Martin, 1996, 1999.
\textsuperscript{1553} For further analysis see Kuusela, forthcoming.
\textsuperscript{1554} Davidson, 1987.
\textsuperscript{1555} Davidson, 1991, 163.
limit we place on the social sciences is set by us when we decide to view humans as rational agents with goals and purposes that are subject to moral evaluation. This decision has enormous importance. According to Martin we need a view like AM because without an anomaly there is no human being; the human being is itself an anomaly.\textsuperscript{1556} Is it essential for our self-conception, for our psychological well-being, that we see ourselves as anomalies of nature? I find this line of reasoning very tempting. It seems to me that Wittgenstein, Davidson and von Wright stress this idea in their different ways. Wittgenstein wrote that: "[…] unforeseeability must be an essential property of the mental."\textsuperscript{1557} Davidson concluded: "When we are making up our minds what to do or what to think we cannot at the same time conceive of our reasoning as bound by the strict laws of physics."\textsuperscript{1558} Finally, von Wright suggested: "The conceptual basis of action, one could say, is partly our ignorance (unawareness) of the operation of causes.…"\textsuperscript{1559} This is part of the fascination of being human; we err and are not transparent to ourselves; we are surprised to find that we act against our better judgment or that others do not act as we have predicted. Perhaps in the future a physicalistic conception of a human being could factor out the anomalies and inadequacies of human life. In this process it would factor out also the human being. Should we be daring enough to allow the dramatic re-shaping of our-self conception? Von Wright thought that the future life of “humans” who would merely behave would appear incomprehensible to us, but if this is our fate we should not be fearful of it – because the fears that are real now could not be expressed any longer, in fact they would simply not exist, in the reshaped form of life. To me this sounds like a suggestion that if future generations are born in prison, it does not worry them because they do not know and cannot dream of a different reality. When faced with the enthusiasm of technology and its promise that all problems of human life will be solved, we can decide now how much we value our current self-conception and whether we want to preserve it.

Incidentally, Martin started to appreciate AM’s political dimension when he heard Patricia Churchland suggesting that certain people should be “fixed”. This was ten years ago. Neurophilosophers who are now “at work” suggest that in the future RAM-chips should be installed in peoples’ heads and that a brain-based philosophy of life should be adopted. Martin sees Davidson’s position as an alternative to these kinds of approaches; it is an alternative way to think about the mental lives of humans. He is not the only one to make this

\textsuperscript{1556} Martin, 1996, 1999.  
\textsuperscript{1557} Wittgenstein, 1992, 65e.  
\textsuperscript{1558} Davidson, 1999u, 359.  
\textsuperscript{1559} Von Wright, 1970, 30.
interpretation. Ramberg has claimed that: “Scientism... renders us subject to certain forms of oppression”\(^\text{1560}\) and concludes that Davidson’s views can be seen to provide tools “[...] in a struggle against the steady spread of the dehumanizing, homogenizing management of human existence that is the real threat of scientism.”\(^\text{1561}\) These tools may help us to challenge the idea of humans which seems reasonable now when it is so blindly believed that a scientific discovery is always an improvement.

Great scientific improvements can be used also as tools for great oppression; the techniques which would allow causal explanations to reign over mental explanations could be used to challenge the autonomy of the individual. On the hand one, causal explanations could be used to disqualify the verdict of the acting agent about his efficacious reasons. On the other hand, these explanations could be used to justify behavior which we would otherwise judge as unacceptable. These are the worst-case scenarios. If we decide that, as suggested in section 4.2.2, people are less eager to help others when they are a part of a group or that referees unconsciously favor athletes dressed in red, shouldn’t these findings have consequences for the ways we morally evaluate the people in question? If a referee treats an athlete unfairly, can we blame him if his visual system is biased and his adaptive unconscious makes the “decisions”? If in these cases we decide to withdraw our moral evaluations, this is also a statement which downgrades the autonomy of the individual, because we are expressing an attitude that a human being is completely at the mercy of forces which are unknown to him.

Wittgenstein was a lone wolf. Davidson, despite the originality of AM, stayed quite faithfully inside the tradition of analytic philosophy. Von Wright’s position differs from those of Davidson and Wittgenstein because he was very concerned of societal problems, participated in public discussions and attempted to reach a wider audience through his non-philosophical writings.\(^\text{1562}\) As a philosopher von Wright had a side which was lacking in Davidson and Wittgenstein, and it is this side which best explains von Wright’s motivation for adopting a non-reductive position.\(^\text{1563}\) My claim is that von Wright’s philosophy of mind can be correctly understood only against the skeptical attitude that he had towards his times.

\(^{1560}\) Ramberg, 2000, 367.

\(^{1561}\) Ibid.


\(^{1563}\) Whereas von Wright’s writings meant mainly for professional philosophers contain subtle arguments which are often put forward hesitatingly as modest suggestions, many of his writings for a wider audience do not conceal the opinions of the author. There is an interesting difference in the styles in which these separate areas of von Wright’s philosophy are written. Von Wright was never quite clear about how the two philosophers relate to each other. Egidi, 1999a, 2009, has explored the connections of von Wright’s “two philosophies” in an interesting way. For a brief general discussion of these connections see Stoutland, 2009.
Von Wright’s broader concerns were at the forefront of his work during the 1980s, but the implications of scientific progress worried him until the end. He was troubled by the “apocalyptic mood” which cast its shadow over the end of the 20th century, and even noted that his pessimism had given rise to a kind of depression. I think that von Wright’s remark about Wittgenstein applies also to his own position:

Wittgenstein… thought… that the problems with which he was struggling were somehow connected with the ‘way people live’… with features of our culture and civilization to which he and his pupils belonged. His attitude to this culture was… one of censure and even disgust. He therefore wished these ways of life changed, but he had no faith that he or his teaching would change them.\[1564\]

Von Wright was perhaps too much of a humanist to describe his own view towards society in terms of “disgust”. But, so it seems to me, he was deeply disappointed, or perhaps saddened is the right word, to see where scientific rationality had brought humankind. Humans should be able to reason about what is best for them. The current situation, however, shows that we have not been able to use our wisdom wisely. Many of the most severe problems of today are a result of the arrogant belief that humans can control all the outcomes of rationality.

Von Wright certainly thought that there were many things wrong with our current way of life. It is interesting how different conclusions can be drawn from the triumph of science. I have briefly described how Paul Churchland welcomes technological innovations because they contribute to the well being of mankind, and how he wants to come up with a technologically more powerful theory of human nature which enables people to stick RAM-chips into their brains and enhance their cognition. Von Wright does not share this vision. Whereas Churchland is eager to point out the alleged benefits of many technological luxuries, von Wright did not hesitate to stress the problems that technology has brought along with it. Private motoring and mass-tourism he described as “plagues”, noting that he did not want to “sugarcoat” things by using euphemisms as many are willing to do in our fraudulent world.\[1565\] Von Wright’s conviction was that if the current state of the world and the reasons for it were properly considered, a suggestion that the standards of living or material well-being should be increased would sound so hollow and absurd that anyone with even moderate intelligence would be embarrassed to use these suggestions in the debate concerning the nature of future society.\[1566\]

\[1564\] Von Wright, 1982, 110.
\[1565\] See von Wright, 1997b.
\[1566\] Von Wright, 1987.
Believers in progress, like Churchland or Gazzaniga claim without hesitation that we are better off than before. Von Wright thinks that this comment is both stupid and thoughtless, because usually the term “we” refers to Western people, ignoring the majority of people living today. Von Wright concluded that a typical Western country is “an island of happiness” in a troubled word, but it is unclear whether increased purchasing power and other merits of modern society show that the people of these societies are psychologically healthier, more satisfied with their lives or better oriented to the conditions of living than previous generations. The rationalism inherent in Western culture has created a fragmented picture and with the loss of unity, “[...] the world around us assumes increasingly absurdist features. Life loses its meaning. Our rationality turns out to be only the reverse side of a new form of irrationality.”

Von Wright predicted that in the future an increasing number of people would find their labor meaningless and for them society would seem inhuman; the result would be massive alienation. An individual human being cannot really comprehend the overall situation. The possibilities of participating in decision making processes diminish. This widens the gulf between the power elite and the masses of vassals. Unemployed and other unfortunate people who have been pushed aside are in danger of becoming “Untermenschen” who remain passive and adjust to the contours of life dictated by the dynamic and effective specialists of economy, science, technology and administration. Von Wright concluded that this bifurcation creates a substrate for attitudes which may be truly described as fascist.

Von Wright’s view of the current state of the Western civilization was very bleak, and to him the future did not look any brighter. Together with politics, philosophy he saw as being in a state of helplessness, which revealed itself in the ways that philosophers were clinging on to formal logic or turning their interests to the history of philosophy; these trends in philosophy von Wright saw as forms of escapism. Many philosophers do not address the severe problems of our times. But should they face them? In chapter one, I claimed that philosophy is distancing itself from ordinary life; perhaps this work and many of the questions discussed in it testify well to this claim. It is a telling fact about the current state of philosophy that many philosophers have either allied with science or have constructed a self-sustaining discussion, for example, about the modal nature of supervenience. I also suggested that in times like ours it should perhaps be a responsibility of a philosopher to take a stance on the practical problems facing humanity, which are so severe. Perhaps a philosopher should feel

---

1567 Von Wright, 1987a.
1568 Von Wright, UPj, 8.
special responsibility towards others. This is of course only a personal opinion and I cannot
convince the reader that this is what every philosopher should do. But I agree with von
Wright’s plea:

When the system of values that underlies a culture weakens, those values may become false. Then the
main duty of a philosopher is… to philosophise with a hammer. At such a time the philosopher must
break and destroy the old… the philosopher will display a kind of anarchy… as a precondition for a
new, healthier culture. Our own time is such a time. My own belief is that the best thing philosopher can
do today is to act as a critic of his own time.\footnote{In an interview with Saarinen, 1986, 129.}

Perhaps we are already doomed, but von Wright saw two alternatives as far as the future was
concerned, and only time will tell which will actualize. On the one hand, we may learn to
tolerate the conditions of life which seem utterly inhumane to us now. We may come to
accept that humans are navel-gazing consumers and nothing more. But for us to whom this
consequence and conception of humanity is intolerable, von Wright’s hesitant final alternative
may seem to be worth developing and exploring. He notes that his hope, insofar as one can
speak of such, is directed towards the protest which arises against this so-called “progress”
from within. This possible protest is grounded in the same power which has created the
current situation: human rationality.

At this point the reader is certainly wondering how cultural criticism can possibly
relate to the irreducibility of the mental. In chapter one I raised the question of what a
philosopher’s motivation for defending a certain position could be. Positions in the
philosophy of mind do not exist in isolation and they often indirectly express a commitment
to a certain kind of ideology or to certain values. In many cases the motivation for defending a
certain philosophical view can be traced back to the worldview of a philosopher. We should
not doubt that the proponents of scientific philosophy of mind honestly think, or hope, that
scientific progress will benefit, if not mankind, at least a the fortunate Western people that
comprise part of it. But I believe that those, like Wittgenstein, Davidson, and von Wright who
argue for the irreducibility of the mental believe just as firmly that the threats posed by
technological progress are greater than the possible benefits. Since von Wright thought that
Western civilization was in decline and that we are living in apocalyptic times, it is plausible
to think that this kind of all-encompassing view influenced his specific philosophical views.
He proposed, in a very different manner than the admirers of progress, that: “[...] we have
arrived at a station when something has come to an end and when reconsideration of our
destiny is imperative. Spiritually, we are in a period of what I propose to call \textit{reflective dusk}.

\footnote{In an interview with Saarinen, 1986, 129.}
Before us is the impenetrable darkness of night.”¹⁵⁷⁰ There cannot be any proof that this is so, but it is possible to see things from this perspective. From this point of view the new trends, like a thoroughly material mind, may appear as signs of regression and not as steps forward. It is telling that whereas Paul Churchland thinks that the progress of technology will turn us into beings who communicate with “Übersetzen”, von Wright thinks that the same progress will turn the majority of people into “Untermenschen” who feel alienated and disconnected from reality.

Whereas Churchland wants to escape the prison of fate through technology, by controlling, manipulating and modulating the course of nature, von Wright humbly proposed that humans should recognize and understand their limits; we are masters and slaves of our own destiny. This is what it means to be human. It is not a negative remark about human nature, but a comment on how the human condition can be understood. The lives of humans show that there is no human nature the essence of which could be found from the brain, and consequently there is no essence of humanity which could be shaped according to an ideal based on science. There are only individual human beings who act in imperfect and sometimes hideous ways when trying to shape their lives according to their own understandings of what is valuable and what is not.

Wittgenstein feared science could lead us to think that our ordinary ways of understanding others are inferior when compared to sophisticated scientific methods. He thought that extraordinary scientific achievements could be used to destroy the “human soul.” It is therefore understandable why he was not enthusiastic about a scientific philosophy of mind. Davidson, as we have seen, suggested that the limit of reducibility of the social sciences is set by us when we decide to view men as rational agents with goals and purposes who are subject to moral evaluation. It is revealing that Davidson begins one of his earliest articles by posing the following question: “Can we hope to give an exhaustive description of man and his behaviour, in the physical vocabulary or concepts of the physical sciences, that is, the sciences which treat of non-animal, non-human phenomena on a par with others?”¹⁵⁷¹ He notes that we wish to make clear where we stand with respect to this question and then goes on: “[…] do we have any choice, but to stand where the truth is? […] I believe that there is something quite appropriate in the idea of taking a stand in this matter. There is an element of choice involved in the answer to this question, in my opinion.”¹⁵⁷² Why should the choice be

¹⁵⁷⁰ Von Wright, 1997d, 13.
¹⁵⁷¹ Davidson, 1964, 226.
¹⁵⁷² Ibid.
made? Davidson is not worried about the possibility that we could invent thinking machines, conscious robots to which mental vocabulary could be meaningfully applied.\textsuperscript{1573} But he is worried about the possibility that “[...] at some point we should stop applying concepts of rationality, intentionality and moral evaluation to men.”\textsuperscript{1574} I suggest that this worry is ultimately Davidson’s reason for defending the irreducibility of the mental and grounds the motivation for his anti-reductionism.

Martin, who suggested that AM is an ethical-political position wished that “[...] Davidson would be a bit more explicit about what makes him, in the general context of analytic philosophy, a real humanist, especially in comparison to the ‘neurophilosophers’ and others who are attempting to show that humans are just a more complex form of thermostat or frog.”\textsuperscript{1575} Here is my answer to Martin. At the very beginning of his career, Davidson thought that we should take stand in this matter because not taking a stance could lead to a point where rationality, intentionality and morality would not be part of the human life. I suggest that when Davidson – forty years after taking the stand – is still willing to resist the “irresistible urge” to consolidate our vocabularies of explanation, it is the same motivation which grounds his will. A possible consequence of a scientific philosophy of mind would be the dehumanization and mechanization of humans because we would no longer see humans as rational and moral beings, but more like physical objects functioning according to physical laws. The behavior of these objects would be freed from any “anomalies” or “residues of meaning”. But if we think that the human being is an anomaly, then a complete scientific understanding of the nature of humanity would also be the end of humanity. I am not sure that there can be an argument demonstrating that this result must be prevented at all costs; the decision to argue against a scientific understanding expresses no more, and no less, than a commitment to what we find important.

Unforeseeability belongs to the human condition. If we say that we cannot figure a person out, this is not a similar kind of remark to the comment that we cannot figure out a certain mechanism. This is one essential distinction between a human and a machine. Suppose that we could predict perfectly the behavior of others. This would diminish the difference between humans and machines and Wittgenstein, like Davidson, feared that it would eventually change our attitudes towards other humans. We all know the basic facts of human life; we know that those who cry are in pain and not pretending, we feel empathy when we

\textsuperscript{1573} This possibility, which did not worry Davidson in 1964, has for many philosophers and scientists become increasingly worrying in the 21\textsuperscript{st} century.
\textsuperscript{1574} Davidson, 1964, 231.
\textsuperscript{1575} Martin, 1996, 120.
see pain in the eyes of others, we treat each other as persons – at least from time to time. But as Wittgenstein: suggests:

[… ] imagine that a kind of thermometer is used to ascertain whether somebody is in ‘pain’. If someone screams or groans, then they insert the thermometer and only when the gauge reaches this or that point do they begin to feel sorry for the suffering person, and treat him as we do someone who ‘obviously is in pain.’

We have not yet reached this position in our society and the thought-experiment may seem queer. If we see that somebody is in pain, why use a special thermometer? He who has need for such a device is still confused about the symptoms and criteria of mental phenomena. Those who would feel sorry for others only after consulting their thermometers would be strangers in a society which is formed in accordance with our current self-conception. As Wittgenstein notes: “Where measuring is not important we don’t measure, even if we are able to.”

But on the other hand, there are signs that the criteria for mental states are changing from behavioral to purely physical. Given the development of fMRI-scanners and other brain-imaging techniques, the case that Wittgenstein asks us to consider may be close to reality. In section 3.2, I noted that scientists are suggesting that “mind-reading” is already possible and articles like “Decoding Mental States from Brain Activity in Humans” , “Reading Hidden Intentions in the Human Brain” and so on are being published. Will the criteria for mental phenomena change, and do we as a result become creatures who feel sorry for others only after checking our thermometers? Wittgenstein asks:

[…] would [we] give up our language-game which rests on ‘imponderable evidence’ and frequently leads to uncertainty, if it were possible to exchange it for a more exact one which by and large have similar consequences. For instance, we could work with a mechanical ‘lie detector’ and redefine a lie as that which causes a deflection on the lie detector. So the question is: Would we change our way of living if this or that were provided for us? And how could I answer that?

The question of whether we will change our way of living if these new possibilities are provided is relevant now. And how could we answer that? Do we want to defend a common-sense conception of ourselves or are we willing to hand over to experts the possibility of passing a final verdict about the mental which is based on physical, objective evidence? I claim, at the risk of being “unscientific”, that we have to prevent mental–physical reductions in order to save the purposes of mental concepts and the form of life which is made possible

1576 Wittgenstein, 1992, 93e.
1577 Wittgenstein, 1992, 94e.
1578 Haynes and Rees, 2006.
1579 Haynes et al. 2007.
1580 Wittgenstein, 1992, 95e.
by them. This is what we can call the *inescapability of the* mental. On the one hand, we simply cannot escape the fact that certain behaviors of a human being means for us that he is in a pain; we automatically see the pain. Usually a human being feels sorry for those who are in pain; this is also a fact which we cannot escape. On the other hand, understanding people in terms of mental concepts is mandatory as long as we *want* to see people in a certain way. If we got rid of the mental concepts then our whole attitude towards other people would change; we would not see the consciousness in their faces and our attitude towards others would not be an attitude towards other souls.

Wittgenstein felt disgust for philosophical theories and his critical view about the darkness of his times explains a part of his resistance against a reductive view about the mind. Davidson, as far as I know, was not especially hostile towards the currents of his times, but his view about the mind, which is based on the *choice* to defend non-reductivism, is exceptional in the contemporary philosophy of mind. Von Wright’s position differs in an interesting way from the positions of Wittgenstein and Davidson. He doubted whether the triumph of technology had increased the quality of life and feared that technology might be misused. Von Wright was clearly worried that technological developments allowed for the increased manipulation and control of human beings and decreased the possibility of real democracy and freedom. For von Wright, the problem of human freedom was especially a *social* problem and the more metaphysical aspects of this problem were secondary. The plausible consequences of technology for human freedom and autonomy von Wright saw as negative. Man’s emancipation and autonomy as an independent agent were in danger of becoming lost because the global system based on technology was continually battering the world of experience.

Given von Wright’s worries about the negative consequences of technological progress could he really have defended a reductive view about the mind? Could he have defended a technologically-based material mind which could have negative consequences for human freedom? If von Wright believed, *pace* Gazzaniga for example, that technology *will* be misused and that the consequences *will* be harmful, it would have been irrational for him to defend a position in the philosophy of mind which could facilitate a move in this direction. I think that even if von Wright thought that a thoroughly material view about a human being was possible, he could not have defended such a view. Is it a co-incidence that both Wittgenstein and von Wright, deeply pessimistic about their times, ended up defending positions which so strongly emphasized the autonomy and irreducibility of the mental? Is it only a coincidence that those thinkers, like Churchland or Gazzaniga, who have a positive
view about scientific progress and believe that science can solve all the problems it creates, end up defending a reductive view about the mind? Von Wright thought that we should seriously consider what kind of society we wish to build and this question involves a value judgment; the answer is partly subjective, a matter of taste and depends on subject’s aspiration. My claim is thus that the motivation to defend the irreducibility of the mental is essentially related to the choice regarding the kind of society towards whose actualization we should work. The majority of von Wright’s work in the philosophy of mind was done after the idea of progress became problematic for him and after he began to question the foundations of our current lifestyle. In 1979, von Wright speculated that perhaps something radically new against the current Zeitgeist may rise from “hermeneutic understanding”. Roughly twenty years later he suggested that the search for a congenial worldview may lead to a situation where we willingly reject the idea of a complete explanation. Then we would no longer yearn for a general theory encompassing all natural phenomena, we would expect that the phenomena studied by the human sciences would not reduce to phenomena studied by the natural sciences, and this would lead to the rejection of reductions. Von Wright concluded that this result could be called the “deconstruction of Western cultural heritage”, and the consequence would be a diversified following of the scientific world view, a form of pluralism. I think that von Wright managed to reach this position in his philosophy of mind; he chose to reject the tempting idea of an all encompassing view, although he believed that a science with a holistic methodology would for a long time receive less support than natural science.

Are these worries about the progress of technology, about the threat of a material mind, reasonable? Is neo-luddism a gratuitous worry and should a right-minded modern person go along with the technophiles? The development of neuroscience has been fast in the first decade of the 21st century. What literally would have been wild science fiction a few years ago is becoming reality today. Different sciences are being invaded from below by the physico-chemical approach to the study of nature. The possibility of the genetic manipulability of life is greatly influencing biology. Nowhere is our desire to steer the course of nature more evident than in the approaches of the most recent technologies. It is possible to get a glimpse of this invasion from below and the problems relating to it by considering the topics discussed in Neuroethics, which is a journal published since 2009. They include,  

1581 In the interview by Saarinen, 1986, 132.  
1582 Von Wright, 1979.  
1583 Von Wright, 1997c.
among others, “Advancing Neuroregenerative Medicine”\textsuperscript{1584}, “The Future of Psychopharmacological Enhancements”\textsuperscript{1585}, “Knocking Out Pain in Livestock”\textsuperscript{1586} and “Neuroenhancement of Love and Marriage: The Chemicals Between Us”\textsuperscript{1587}. The key concept in this discussion is enhancement. It is emblematic of the 21\textsuperscript{st} century that there is an increasing desire to enhance human cognition by sticking artificial modules into people’s brains or by doping humans with psychotropic drugs – when at time same time the current problems created by technology are left without solution. Most of the desired enhancements are unnecessary from a reasonable point of view; healthy college students use methylphenidate to improve alertness, concentration, and academic performance.\textsuperscript{1588} Many see this not as a form of drug use, but as voluntary self-improvement. Although we may grant people the right to enhance their cognition up to a point, how should society react when students start demanding cosmetic neurology, cognitive prostates or psychotropic drugs? In a society where technological progress is praised, where enhancement has become a norm, it becomes extremely difficult to explain why such cognitive enhancements do not belong to anyone. Should every parent be allowed to enhance their perfectly normal child? Societies cannot promise enhancements to everyone; currently even normal health care is beyond the reach of huge numbers of people. It is utterly clear that the most recent technological innovations will always benefit only those who are already well off. RAM-prosthetics enable the rich and fortunate to enhance their cognition, and this will further widen the gap between the enhanced and non-enhanced. If the neuroenhancement of love is possible, this will never be available for the masses. On the other hand, pharmacological or neuroscientific interventions can be used to standardize the mental economies of the abnormal. Is this the kind of society we want to build and leave for the future generations?

People anxiously want to get rid of negative feelings because we are living in a culture where negative feelings are considered unnatural. Nowadays medicines are the standard device for normalization. The political philosopher John Gray has written that the humanist successors of religion who believe in salvation through science hope that “in the near future everyone can be happy. Societies founded on a faith of progress cannot admit the normal unhappiness of human life.”\textsuperscript{1589} Indeed, Paul Churchland thinks that increased capabilities to control, intervene and modulate the mental lives of humans will “[…] lead to

\textsuperscript{1584} Grunwell, et al. 2009.
\textsuperscript{1585} Schermer et al. 2009.
\textsuperscript{1586} Shriver, 2009.
\textsuperscript{1587} Savulescu and Sandberg, 2009.
\textsuperscript{1588} For a discussion, see Forlini and Racine, 2009.
\textsuperscript{1589} Gray, 2002, 142.
greater humanization of the individual humans… it will allow us to be more kind, more insightful, more caring about other people and much more effective…. I don’t fear it at all on the whole.\textsuperscript{1590} Perhaps this is true, only time will tell. But what prevents us now from being more kind and caring about other people? Should we, instead of developing cognitive prosthetics, consider what we could give up in order to create some righteous balance in the divided world? To give up some of the technological innovations of which we already enjoy and stop the development of new ones is difficult because technology speeds up the rhythm of life so much that the soul of a person cannot sustain the result. Perhaps we will soon need, and not merely desire, cognitive enhancements in order to survive. Many senior citizens are already feeling that they do not belong in a world the progress of which they cannot comprehend. Perhaps science will make us all happy and super-effective; then it would have succeeded in destroying the human being.

The threats arising from reductive views about mental causation are similar to the threats of the material mind. I claim that even if a reductive explanation of mental causation were possible, such an explanation should not be given. Advances in genetic technology have raised the possibility of controlling the bodily and mental development of the individual. A reductive solution to the problem of mental causation would allow the increased manipulation and control of, and intervention in the mental life of a person. But people should not be treated in this way; we should not think of each other as mere brains that can be manipulated. If person’s own judgment is constantly overridden by experts who work in the name of authority, the result is likely a feeling of alienation and this is a direction which we should not take; it is not the way how people should be treated. This is why I argued in section 4.2.2 that we should let precursory reasons override non-mental explanations. We should attempt to come up with a view of how non-physical mental causation is possible; a model referring to precursory reasons is a suggestion about the kind of direction in which we could proceed. We should have more faith in human beings than we currently do and ultimately grant them autonomy by deciding to respect their self-understanding. A thorough self-monitoring could lead to an increased self-understanding, and through that to a more meaningful life. On the other hand, a decision to trust a non-mental cause over a precursory reason would be an expression of a certain kind of attitude towards human beings. Is this attitude objectionable?

Is it possible to provide an argument demonstrating that people should not be exploited? Here I shall not make an attempt at such a difficult task.

In a culture which is based on hyper-rational attitudes and its technological achievements, other forms of spiritual life tend to wither away because they are taken to be inferior. Those who believe in the Great Idea of Progress think that the future will be better than the present. They believe that perceiving the “darker” side of civilizations helps us to move on to higher levels of enlightenment; we do not want to be forced to have bags over our heads. Paul Churchland’s opinion is that: “[...] life now has more meaning, more potential for humans than it has ever had before.”  

This is thanks to science. But I would claim, science the deliverer of mankind is also the standardizer of mankind. Science makes the work of societal judges easier, and perhaps science will eventually lay down the standards of normality. I believe it was this kind of possibility which troubled Wittgenstein when he wrote: “Madness need not be regarded as an illness. Why shouldn’t it be seen as a sudden – more or less sudden – change of character? Why shouldn’t a man suddenly become much more mistrustful towards others? Why not much more withdrawn? Or devoid of love?”

What could set the limits of reasonable human behavior, what could dictate what a reasonable life should look like? Science cannot do this. Von Wright suggested that the only hope is that we make a deep revision of current values and priorities by acquiring new wisdom “[...] which will make us more humble and respectful, not only in our attitude to other cultures, but also to that which is the common frame of all cultures, viz. nature, the Mother Gaia, which nourishes us materially.”

How could this revision of values happen? The restriction of the pursuit of knowledge would be a drastic solution, but if the quest for knowledge threatens our existence then the question of whether the rights of the individual should be narrowed down is perhaps mandatory. It is an unfortunate paradox that the threats of technology may lead to a situation where we have to limit individual freedom; wouldn’t this be another form of fascism which merely replaces the imagined fascism of science and technology? We can only hope that humans are wise enough to willingly abandon destructive projects, but whether this can happen as long as we are driven by an economic system which encourages competition is uncertain.

The culmination of philosophical progress is the achievement of wisdom. Could philosophers help people to acquire the new kind of wisdom which von Wright called for?

---


1592 Wittgenstein, 1980a, 54e.

1593 Von Wright, 1997d, 17.
Insofar as the views of a philosopher influence the thinking of others, a philosopher influences the way of life of others. By changing the attitudes of humans, by changing their lives, a philosopher is changing the nature of reality. The task of philosophy, the task of philosopher, which was considered in chapter one, demands that a protest must be raised whenever lies are being imposed on us and whenever we are asked to forget or accept the injustices that are being carried out. I believe, however, that a radical infiltration of the lives of people cannot be our ultimate savior. In the end, the conclusion which is both edifying and troubling is that humans must change themselves, because whatever evils technology may bring, it is humans using and developing technology which are the real threat. An authority, be it a philosophical, political or scientific, cannot bring real change. In one of his last writings von Wright suggested that: “A restoration of morality can happen so to speak only ‘from below’ in the form of reinvigorated basis of morality in custom and good manners.”

How to change humans then? Wittgenstein wrote:

The sickness of a time is cured by an alteration in the mode of life of human beings, and it was possible for the sickness of philosophical problems to get cured only through a changed mode of thought and life, not through a medicine invented by an individual. - Suppose the use of motor-cars produces or encourages certain illnesses, and mankind is plagued by such illness until, from some cause or other, as the result of some development of other, it abandons the habit of driving.

There is no final answer to the question of how the thoughts and lives of human beings could change for the better; it requires a real change in our lifestyle, an infiltration from below which must start in individuals. Davidson once wrote that it is not easy to act: “when we know a choice open to us will profoundly and irreversibly change our lives and ultimately affect the sort of person we are.” But when we do decide to act: “[…] on the basis of a reasoned consideration of the values involved, and of what will happen to our future selves, this over-viewing of ourselves and our place in the world is perhaps the highest exercise of rationality.”

We can only hope that through the highest exercise of the characteristic human skill, rationality, we can shape our values from below, hold on to the integrity of our souls, recognize the value of other people and abandon the desire to base our humanity on a technologically more powerful theory of human nature. How likely such an outcome is I shall not predict, but as it is sometimes said, today’s dreams are the realities of tomorrow.

---

1594 Von Wright, UPk, 6.
1595 Wittgenstein, 1956, 57.
1596 Davidson, 1999v, 423.
1597 Ibid.
Bibliography


Antony, Louise, 1999, “Making Room for the Mental”, Philosophical Studies 95, 37-44.


Anttiroikko, Ari-Veikko, 1997, ”Käsitteet ja Ilmiöt – Keskustelua G.H. von Wrightin Filosofian Sisäisistä Jännitteistä”, Königsberg 1, 7-21


582


Davidson, Donald, 1967a, “Causal Relations”, in Davidson, 2001, 149-163.


Davidson, Donald, 1973a, “Radical Interpretation”, in Davidson, 2001a, 125-141.

Davidson, Donald, 1973b, “Freedom to Act”, in Davidson, 2001a, 63-83.


Davidson, Donald, 1974a, “Belief and the Basis of Meaning”, in Davidson, 2001a, 141-155.

Davidson, Donald, 1974b, “Psychology as Philosophy, in Davidson, 2001, 229-239.

Davidson, Donald, 1974c, “Replies to Lewis and Quine”, in Davidson, 2001a, 281-287.

Davidson, Donald, 1975, “Thought and Talk”, in Davidson, 2001a, 155-171.


586
Davidson, Donald, 1984, Expressing Evaluations, The Lindley Lecture (monograph), University of Kansas.
Davidson, Donald, 1985b, “A New Basis for Decision Theory”, Theory and Decision 18, 87-98.
Davidson, Donald, 1987, “Knowing One’s Own Mind”, in Davidson, 2001b, 15-39.
Davidson, Donald, 1990a, “Epistemology Externalized”, in Davidson, 2001b, 193-204.

Davidson, Donald, 1993e, “Reply to Peter Bieri”, in Stoecker (ed.), 311-315.


Davidson, Donald, 1994, “Radical Interpretation Interpreted”, Philosophical Perspectives, 8, Logic and Language, 121-128.

Davidson, Donald, 1994a, “Entry on Donald Davidson”, in Guttenplan (ed.), 231-236.


Davidson, Donald, 1997a, “The Emergence of Thought”, in Davidson, 2001b, 123-135.


Davidson, Donald, 1999b, “Reply to Anita Avramides”, in Zeglen (ed.), 155-156.

Davidson, Donald, 1999c, “Reply to Thomas Nagel”, in Hahn (ed.), 207-211


Davidson, Donald, 1999e, “General Comments”, in Zeglen (ed.), 157-161.

Davidson, Donald, 1999f, “Reply to Tyler Burge”, in Hahn (ed.), 251-254.

Davidson, Donald, 1999g, “Reply to Deborah Hansen Soles”, in Hahn (ed.), 330-332.

Davidson, Donald, 1999i, “Reply to Carol Rovane”, in Hahn (ed.), 480-481.


Davidson, Donald, 1999k, “Reply to Pablo Quintanilla”, in Caorsi (ed.), 326-328.


Davidson, Donald, 1999m, “Reply to Jack S. Crumley II”, in Caorsi (ed.), 330-332.


Davidson, Donald, 1999s, “Reply to Herman Cappelen and Ernie Lepore”, in Zeglen (ed.), 100-101.

Davidson, Donald, 1999t, “Reply to Paolo Astiazaran”, in Caorsi (ed.) 335-336.


Davidson, Donald, 2000, “Perils and Pleasures of Interpretation”, *Ars Interpretandi* 5, 22-37.


Dennett, Daniel, 2007, “Philosophy as Naïve Anthropology: Comment on Bennett and Hacker” in Bennett (et al.), 73-95.


Dupre, John, 1988, “Materialism, Physicalism and Scientism”, *Philosophical Topics* 16, 31-57.


Egidi, Rosaria, 1999a, “Von Wright and ‘Dante’s Dream’: Stages in a Philosophical Pilgrim’s Progress”, in Egidi (ed.), 1-35.


Emiliani, Alberto, UP, “Three Challenges for Bi-Polar Monism”, unpublished manuscript.


Gluer, Kathrin, 1995, “Relations and Transitions – An Interview with Donald Davidson”, Dialectica 49, 75-86.


Haynes, John-Dylan et al., 2007, “Reading Hidden Intentions in the Human Brain”, Current Biology 17, 323-328.


Jacob, Pierre, 2002, “To Be or Not to Be In Descartes’ Shadow”, retrieved 27.1.2010 from http://jeannicod.ccsd.cnrs.fr/docs/00/05/33/99/HTML/index.html


Lewis, David, 1974, “Radical Interpretation”, *Synthese* 23;


Mclauhlin, Brian, 1989, “Type Epiphenomenalism, Type Dualism and the Causal Priority of the Physical”, *Philosophical Perspectives 3*, 109-135.


Moser, Paul and Yandell, David, 2000 “Farewell to Philosophical Naturalism”, in Craig and Moreland (eds.), 2000, 4-23.


Niiniluoto, Ilkka, 1990, Maailma, Minä ja Kulttuuri, Otava, Helsinki.


Nisbett, Richard and Wilson, Timothy, “Telling More Than We Can know: Verbal Reports on Mental Processes”, Psychological Review, 84, 231-259


Owen, Adrian, Schiff, Nicholas and Laureys, Steven, 2009, “A New Era of Coma and Consciousness Science”, Progress in Brain Research 177, 399-411.


Quine, Willard Van Orman, 1953a, From a Logical Point of View, Harvard University Press, Cambridge, MA.


Reinholds, Lars, 2006, Interpretation and Explanation: Reflections on Donald Davidson’s Philosophy of Mind, University of Oslo Unipub, Oslo.


602


Stoutland, Frederick, 1999, “Intentionalists and Davidson on Rational Explanation”, in Meggle, 1999, 191-211.


Stoutland, Frederick, forthcoming, “Davidson on Rational Explanation: a Reconsideration”


Wright, Georg Henrik von, 1981a, “Itsekkyys ja Solidaarisuus”, *Näköpiiri* 4, 18-19


Wright, Georg Henrik von, UPa, “In Defense of Psychology”, unpublished manuscript.


Wright, Georg Henrik von, UPc, “Upper Circle – Lower Circle”, unpublished manuscript.

Wright, Georg Henrik von, UPd, “Psycho-physical Interaction and the Closedness of the Physical World-Order”, unpublished manuscript.

Wright, Georg Henrik von, UPe, “Consciousness”, unpublished manuscript.

Wright, Georg Henrik von, UPf, “Supplementary Notes to ‘On Mind and Matter”, unpublished manuscript.


Wright, Georg Henrik von, UPj, “Nicholas Berdyaev’s Philosophy of Technology”, unpublished manuscript. A revised version is printed as von Wright, 2000c.
Wright, Georg Henrik von, UPk, “Custom – The Basis of Morality”, unpublished manuscript.


