Phenomenal Consciousness as a Structure in Our Brains

Department of Philosophy, History and Art Studies
University of Helsinki
Finland

Kristjan Loorits

Phenomenal Consciousness as a Structure in Our Brains

ACADEMIC DISSERTATION

Academic dissertation to be publicly discussed, by due permission of the Faculty of Arts at the University of Helsinki in auditorium XIV, University Main Building (Unioninkatu 34), on the 25th of February, 2019 at 12 o'clock.

ISBN 978-951-51-4868-1 (paperback)
ISBN 978-951-51-4869-8 (PDF)
http://ethesis.helsinki.fi
Unigrafia
Helsinki 2019

Supervised by University lecturer Paavo Pylkkänen
Department of Philosophy, History, Culture and Art Studies
University of Helsinki

Professor Matti Sintonen
Department of Philosophy, History, Culture and Art Studies
University of Helsinki

Reviewed by Professor Antti Revonsuo
Department of Cognitive Neuroscience and Philosophy, School of Bioscience, University of Skovde, Sweden
Professor of Psychology, Department of Psychology, University of Turku, Finland

Professor William Seager
Department of Philosophy
University of Toronto Scarborough

Publicly examined by Professor Antti Revonsuo
Department of Cognitive Neuroscience and Philosophy, School of Bioscience, University of Skovde, Sweden
Professor of Psychology, Department of Psychology, University of Turku, Finland

Professor William Seager
Department of Philosophy
University of Toronto Scarborough

Abstract

Understanding and explaining consciousness has proven to be one of the hardest tasks for contemporary science and philosophy. Despite the remarkable progress in cognitive neuroscience, there are certain fundamental issues regarding consciousness that seem to evade the neuroscientific approach. First, of course, there is the famous *hard problem*: why should any information processing in our brains *feel* like something to us? Why is there something *it is like to be* an organism that processes information in a certain way? Second, there is an equally famous problem concerning the apparent *privacy* of consciousness. According to a powerful intuition, the content of *my* consciousness is directly accessible only to *me*. The above intuition fits nicely with the neurobiological conception of consciousness: it is reasonable to assume that an organism has privileged access to some of its own neural processes. However, it has been argued that we can only talk about phenomena whose defining properties are known to us from the public realm. Accordingly, if our conscious experiences were entirely private, we could not talk or theorize about them. And conversely, if it is possible to talk and theorize about our conscious experiences, then they cannot be entirely private.

If the hard problem has cast some doubt on the explanatory sufficiency of cognitive neuroscience, then the problem of privacy has convinced many that the brain-bound internalist approach to consciousness is fundamentally misguided. In its place, different versions of externalism have been proposed. According to these, conscious experiences should be understood in terms of an organism's relationship to its socio-physical environment.

The view defended in this dissertation rejects the externalist approach and falls under the internalist paradigm. According to the core thesis of this work, conscious experiences are *fully structural* phenomena that reside in our brains in the form of complex higher-order patterns in neural activity. The structural view of consciousness allows us to tackle both the hard problem and the problem of privacy. Regarding the latter, it can be argued that fully structural phenomena are describable and analyzable in public terms even if those

phenomena themselves are private. For instance, one could describe privately experienced neurobiological structures by referring to public phenomena with similar structures.

The hard problem can be formulated in the structuralist context as follows: although many properties of our conscious experiences are clearly structural, there are some that *seem* to be qualitative and nonstructural. Those seemingly qualitative properties, the so-called *qualia*, are at the very heart of the hard problem, for those are the properties that define and determine *what it is like* to be conscious. According to the view defended in this dissertation, those apparently nonstructural properties are, in fact, fully structural. It is argued, based on the neurobiological theory of Francis Crick and Christof Koch, that qualia are the structures of vast networks of *unconscious associations*, and that those associational structures can be found in our neural processes. Differing from other structural accounts of qualia, the above view implies that with the proper brain-stimulating technology, it should be possible to reveal the structural nature of qualia to the experiencing subject directly.

The position defended in this work puts a lot of weight on the notion of structure. To begin with, it is far from obvious what it means for a phenomenon to be fully structural. For example, one might wonder whether the fully structural phenomena are genuinely real in an objective sense, or are they rather artificial constructs tied to some explanatory goals and methods. In this dissertation, the above kind of metaphysical questions are approached from the perspective of structural realism. It is concluded that some structures exist in the world in an objective sense and that conscious experiences are among those structures.

Table of Contents

Acknowledgements

Since my first encounter with philosophy of mind as a freshman student of theoretical philosophy in 2001, my primary philosophical interests have revolved around consciousness. Over the years, those interests have grown and developed in numerous inspiring courses and seminaries arranged and held by Paavo Pylkkänen, who was also a supervisor of my doctoral thesis, as well as my master thesis before that. Paavo's contagious passion for contemporary science and his naturalistic attitude toward traditional philosophical questions have influenced me profoundly. His encouragement and help with countless research-related issues as well as practical matters have been priceless, and I am most grateful to him for that.

I am also grateful to my other supervisor, Matti Sintonen, whose support and guidance, especially during the early phase of my doctoral studies, have been of vital importance. It was Matti who encouraged me to approach consciousness from the perspective of structural realism, and I am glad I took his advice. I also want to thank both Paavo and Matti for their help in finding research funding, without which this dissertation would have never been finished. For three consecutive years the funding was provided by the Finnish Cultural Foundation and the final year was financed by the University of Helsinki. I own my deepest gratitude to both organizations.

For invaluable comments for different parts of this work, I am grateful to Paavo Pylkkänen, Antti Revonsuo, Gabriel Sandu, Vili Lähteenmäki, Olli Lagerspetz, Hemmo Laiho, Donnchadh Ó Conaill, Jaakko Hirvelä, Pii Telakivi, and Andreas Fjellstad.

I would also like to thank Terhi Kiiskinen for his help with numerous practical matters during the early years of my philosophy studies.

Helsinki, August 2018

Kristjan Loorits

List of Original Publications

- "Structural Qualia: A Solution to The Hard Problem of Consciousness", 2014, Front. Psychol. 5:237. doi: 10.3389/fpsyg.2014.00237.

- "Dreaming about Perceiving: A Challenge for Sensorimotor Enactivism", 2017, Journal of Consciousness Studies, 24 (7-8).

- "The Location and Boundaries of Consciousness: A Structural Realist Approach", 2018, Review of Philosophy and Psychology, 9 (3).

-

Introduction

1. An overview of the general strategy

This dissertation consists of three articles that examine the prospects of the so-called *internalist* approach to *phenomenal consciousness*. The position ultimately reached and defended is a version of token identity theory—a view that all our *phenomenal experiences* are literally inside our heads by being (token) identical with some of our brain processes. A more precise understanding of the notion of *phenomenal* is sought in the following section, but according to the initial idea, phenomenality is an aspect of consciousness that is defined solely by how things *seem* or *feel* to the experiencing subject.

The proposed internalist view is defended by appealing to some recent developments in cognitive neuroscience, philosophy of science, naturalistic metaphysics and empirical dream research. Overall, the dissertation can be viewed as a naturalistic and scientifically motivated approach to certain features of phenomenal consciousness (including the notorious *qualia*) that are often thought to remain outside the reach of the empirical sciences. Approaching phenomenal consciousness from a neuroscientific perspective is not a new strategy, but the novelty of the line I take consists in combining a neurobiological account of qualia (Crick and Koch 1998, 2003; Koch 2004) with a version of structural realism (Ladyman et al. 2007).

The goal and focus of this research is twofold: First, to defend consciousness internalism against some of its main rivals (i.e., different forms of consciousness externalism, focusing mainly on the sensorimotor theory). And second, to put forward a positive, scientifically motivated and genuinely illuminating hypothesis that would address the so-called *hard problem of consciousness* within the confines of the internalist framework: why is there *something it is like to be* conscious—something it is like to *experience* greenness, redness, painfulness, etc.? More specifically, how might we understand this peculiarly ineffable yet specific "somethingness" or "suchness" in terms of neural processes?

The outcome is a view according to which all scientifically accessible real phenomena (objects, properties, processes, events, etc.) can be analyzed in fully structural terms—with phenomenal consciousness being no exception. Such a view is in direct dialogue, as well as in direct opposition, with David Chalmers' well known position according to which traditional scientific methods cannot explain consciousness precisely because they are only suitable for studying, explaining and describing different relational structures (e.g., Chalmers 2003). Chalmers' pessimistic conviction rests on the assumption that consciousness is (or has) something over and above its structure—the very assumption that is questioned and criticized in this dissertation.

Thereby a significant explanatory burden falls on the idea that phenomenal consciousness itself (and not merely its neural basis) is a fully structural phenomenon. And at the end of the day, any account that claims something like that should convince the reader about the fully structural nature of her own consciousness, including the qualia—the apparently primitive and nonstructural *raw feels* of experienced greenness, redness, painfulness, etc. Therefore, a method is introduced that allows one to "peek inside" some of her own qualia and perceive their internal structures. It is suggested further that with the help of a sufficiently capable brain-scanning and stimulating technology, every quale could be "phenomenally dismantled" in a similar manner. Finally, by appealing to the work of Francis Crick and Christof Koch (1998, 2003; Koch 2004), it is hypothesized that the defining structure of any given individual quale is analyzable in both phenomenal and neural terms—thus narrowing (if not closing) the famous explanatory gap between phenomenal consciousness and the brain.

However, when defending an internalist position regarding phenomenal consciousness, a mere positive explanatory attempt is not enough. The very notion of internalism suggests that there are externalist alternatives to be considered. And indeed there are. Generally put, externalist approaches pose two types of threats to consciousness internalism. First, a significant part of nearly every externalist account of consciousness consists of direct criticism of consciousness internalism. Second, many externalist accounts of consciousness seem to enjoy genuine explanatory success. Therefore, the defense of any internalist hypothesis should include two corresponding parts: it should respond to the externalist criticism, and it should explain why some

externalist consciousness theories seem to be explanatorily successful. In addition, an internalist account can obviously criticize the externalist approach directly by pointing out its flaws and weaknesses.

Following the above guidelines, the main tasks of an internalist account of phenomenal consciousness can be summarized as follows: First, the account should effectively counter relevant externalist criticism. Second, the account should make a case against externalism as a viable explanatory alternative. Third, the account should explain why some externalist theories seem to be explanatorily successful. And fourth (and perhaps most importantly), the account should put forward a positive and genuinely illuminating internalist hypothesis regarding the nature of phenomenal consciousness.

Needless to say, the tools and methods to accomplish the above tasks may vary greatly. The gist of the positive hypothesis proposed in this dissertation is already sketched above. The main criticism against consciousness externalism focuses on the existence of certain non-veridical experiences (dreams, hallucinations, etc.) that resist externalist explanations. Thereby the negative strategy follows the lead of Antti Revonsuo, who has argued repeatedly against consciousness externalism on the grounds of the existence of dreams (e.g., 2006, 2015).

The remaining two tasks, countering externalist criticism and explaining why some externalist theories seem to be successful, are approached by using an analytical toolkit borrowed from structural realism. Externalist criticism against consciousness internalism comes in many different varieties, but some of the most persistent and popular accusations have their roots in philosophy of language. For example, it is often argued that while referring to our phenomenal experiences, we are usually referring to some public and external phenomena, and even when referring to the content of our hallucinations or dreams, we are still bound to use the vocabulary of public phenomena. To be sure, we may dream about things that do not exist (e.g., unicorns or goblins), but as far as we can talk or theorize about the content of our dreams at all, we are always talking or theorizing about something that acquires its meaning from the interpersonal public sphere. It is therefore argued that while talking about our conscious experiences, we simply *cannot* be talking about our neural processes, for these are not publicly accessible in most of the situations in

which we can meaningfully talk about (the content of) our conscious experiences.

Structural realism allows us to respond by showing that one can (and should) separate two questions that are entangled in many philosophical debates and arguments: first, where in the world one can find a certain phenomenon (i.e., a certain structure), and second, what kind of vocabulary or language can one use to describe that phenomenon? Put bluntly, in the light of structural realism, it is possible to describe fully internal brain-bound phenomena by using the vocabulary of external and public objects. If the above is true, then the *meanings* of the words we use do not necessarily contribute anything to the *constitution* of the phenomenon they describe (although in some cases they may).

The success of the externalist approach can be partly explained by the following hypothesis: as far as internal phenomenal experiences can be described (or referred to) at all, they can be described only by using the vocabulary of external and public phenomena (at least in the informal context). Or, more simply, we can only talk about our internal and private experiences by describing their *structures* in terms of public and external phenomena. Second, it is consistent with structural realism that the structures of our *veridical* experiences can be found simultaneously in several different locations: inside our brains and in extended systems that include elements of the environment. Therefore, it is consistent to hold that the positive claim of externalism is genuinely true of veridical experiences, but that internalism is true of *all* experiences (i.e., veridical and non-veridical alike). Obviously, if such happens to be the case, then the (positive) truth of externalism about veridical experiences would explain its explanatory success regarding veridical experiences (without threatening the general explanatory superiority of the internalist approach).

It should be acknowledged that arguments from philosophy of language do not exhaust the externalist criticism against consciousness internalism. Some other externalist complaints are considered and addressed in the articles further below.

Since contemporary consciousness research is a complex multidisciplinary undertaking, it is worth considering where this dissertation is situated in its

wide and diverse field. Despite its naturalistic inclinations, the approach taken here is clearly philosophical in two key respects. First, it includes no empirical experiments and thus provides no new empirical data. Second, it focuses almost exclusively on traditional philosophical issues, such as the hard problem of consciousness, the problem of subjectivity, the problem of qualia, the problem of privacy, the internalism-externalism debate, etc. That said, there is a sense in which the approach is empirical in spirit. At the core of its positive part is an empirical prediction that with proper technology, it should be possible to manipulate the brain in a way that reveals the structural nature of qualia to the experiencing subject directly.

2. Background: The big questions

What is consciousness? Besides wide disagreement on the right answer to this question, there is additional disagreement over what is being asked in the first place. In other words, there is a serious lack of consensus on what we *mean* by consciousness whose explanation we are seeking. One of the deepest divisions in that regard runs between externalists and internalists about consciousness.

Generally, internalists proceed from the initial premise that conscious experiences are something private and "within" an experiencing subject—after all, only *I* can experience *my* experiences (although others may, to some extent, infer the content of my experiences indirectly). Admittedly, if an experience is veridical, then it is typically *about* some actual external object in one's environment; however, the experienced object is arguably not a *part* or a *constituent* of the experience itself, but merely an external *cause* of it (e.g., Revonsuo 2006, 2015; Searle 2000; Clark 2009). According to the internalist intuitions, what one *means*, for example, by a visual experience of a tomato is something clearly distinct from the actual tomato the experience concerns. Therefore, when looking for an explanation for the experience itself (be it veridical or not), we should always look for an explanation of something that is private and "inside" an experiencing subject. And proceeding from that premise, the brain and the neural processes in general seem to be strong candidates for the proper explanatory basis.

By contrast, many externalist approaches set greater weight on the fact that the majority (if not all) of our conscious experiences are *about* something. For example, it has been argued that our experiences are defined by their *intentional* or *representational content*, which cannot be understood or explained in terms of intracranial neural processes (e.g., Tye 1995; Dretske 1995; Lycan 1996). Alternatively, it has been argued that the content of our experiences is intimately linked to the ways in which we express ourselves. If so, then the content of an experience of a tomato would depend, if not on the existence of actual tomatoes, at least on the existence of the concept of tomato, whose meaning is determined by its role in our linguistic community (e.g., Dennett 1991; Tye 1995; Dretske 1995; Lagerspetz 2002). Similarly, the currently popular sensorimotor theory holds that what one *means* by saying that she has an experience of a tomato, is that one stands in a certain relationship with an actual tomato, or in the case of non-veridical experiences (e.g., dreams or hallucinations), that one is in a state (e.g., in some kind of dispositional state) that can be described *as if* standing in a certain relationship with a tomato (e.g., being disposed to act *as if* a tomato were nearby; see Beaton 2013; Myin 2016). By and large, according to the above ways of thinking, what we *mean* by conscious experiences is something that involves our relationship with the world.

The above sketches are rough and simplified, but their purpose is merely to illustrate the general idea: some of the most fundamental disagreements between competing consciousness theories are not about the specific underlying mechanisms of consciousness, but about the meaning of the notion of consciousness.

A popular internalist response (perhaps more often tacit than explicit) to the externalist challenge outlined above consists in appealing to the notion of *phenomenality*. The basic idea is this: let us suppose it is true that conscious experiences are embedded in the socio-physical environment in the way suggested by externalists. If that is the case, then in a world without tomatoes and the corresponding concept, one might still experience (e.g., hallucinate or dream) something "tomato-like," but one could not experience tomatoes. But if so, then in what respect exactly are genuine experiences of tomatoes similar to tomato-like experiences that may occur in a tomatoless world? According

to the general internalist understanding, such experiences are similar with respect to their *phenomenality*.

Let us indulge briefly in a thought experiment. Let us imagine we are transporting a subject who is dreaming about a red tomato to a world with no tomatoes or any other red or spherical objects. Now, should the person wake up and try to describe her dream to the locals, she would fail to do so—the content of her dream would *mean nothing* in the alien world. Nevertheless, according to the internalist understanding, the content would still exist and be *phenomenally similar* to the content of dreams about tomatoes here on Earth.

Most consciousness internalists use the notion of phenomenality in precisely that sense. It is assumed that every individual conscious experience has a certain directly felt aspect whose nature is independent of how we choose to call it or by which linguistic mechanisms we can refer to it—and that (and that alone) is the aspect to be studied and explained within the internalist framework. Adopting the above meaning of the notion of phenomenality allows one to hypothesize that certain neural processes are sufficient for the existence of a *phenomenal* experience of a tomato (or a tomato-like phenomenal experience, if you like), even if they are not sufficient for the existence of an experience of a tomato in some broader sense.

However, not everyone agrees that there is an aspect of consciousness that can be separated from its representational or behavioral properties (e.g., Tye 1995; Dretske 1995; Hurley 2010; Dennett 1991). It can be argued that if the phenomenal aspect is something we can talk and theorize about, it too must be assigned a shared and public meaning. And even if a phenomenal content might lack public meaning in some imaginary world, it must always have one in our world—for otherwise we could not talk about it or construct thought experiments about it.

Again, the above sketches are crude and simplified, and they fail to do justice to many important nuances in the multifaceted debate between consciousness internalists and externalists. However, they suffice to draw attention to one of the most central and persistent themes in that debate: the idea that we can describe and thereby understand our conscious experiences *only* in terms of some external and public phenomena. It is safe to claim that the above idea has a significant role to play in every single externalist account of phenomenal

consciousness—from traditional versions of representationalism to the currently popular sensorimotor theory. Regardless if an externalist understands conscious experiences in terms of behavioral tendencies, representational content, organism-environment sensorimotor interactions, or some other environment-involving phenomena, she understands them in the very same terms she *describes* them. Put bluntly, the idea that terms of description may be quite different from the nature of the described is alien to the externalist way of thinking. Therefore, the internalist framework would benefit greatly from a clear and intelligible account of how a phenomenal content can be internal (and in a sense private) even if described in terms of external (and public) phenomena. What follows is an attempt to provide such an account.

The basic idea is fairly simple: it consists in suggesting that our phenomenal experiences are *fully* structural phenomena (in the sense specified below). And *if* our phenomenal experiences are fully structural, *then* they can be described by whichever means we choose, as long as we get their structures right (e.g., the structure of a melody that exists "in my head" can be described by referring to a famous song with a similar structure). In addition, if phenomenal experiences are fully structural, then we are free to hypothesize that they (i.e., their structures) can be found in neural processes.

The above hypothesis is defended as follows: it is first argued that some simple properties of our phenomenal experiences are structural in a clear and easily graspable sense, and that at least some of those structural properties can be also easily found in our brains. It is then speculated that the rest of the structural properties of our phenomenal experiences will be found in our brains eventually—or at least that this is a perfectly legitimate empirical hypothesis. Next, it is acknowledged that some properties of our phenomenal experiences *appear* to be irreducibly qualitative and thereby nonstructural. Those are the properties that give rise to the so-called *hard problem*. Finally, it is argued that those apparently nonstructural properties can be analyzed in structural terms after all, and that their structures can be described and understood in both neural and phenomenal terms.

Let us begin with a simple example: a dream about a slowly flashing red light with a specific rhythm. Let us acknowledge the fact that the dreamed rhythm is a well-defined fully structural property of a phenomenal experience. The

rhythm is fully structural in the sense that it is defined by its one-dimensional temporal structure (different rhythms differ by their temporal structure, and the same temporal structure would count as the same rhythm). And the rhythm is a genuine property of the phenomenal experience in the sense that it would be logically impossible to remove or alter the rhythm without altering the corresponding phenomenal experience.

Now, let us suppose also that the dreamed rhythm is entirely endogenous—nothing in the dreamer's environment or behavior corresponds to it. Let us assume further that the dreamer will never describe or report her dream to anyone. The way the example is constructed, the *only* place in the entire physical universe where one can find the experienced rhythm is in the neural processes of the experiencing subject (of course, a similar rhythm could emerge in some other location accidentally, but the only place where the rhythm can be found *systematically* is in the neural processes of the subject). More specifically, the rhythm can be found by monitoring the subject's neural correlate of redness. Admittedly, things might get more complicated if the experienced rhythm is very rapid, which is why the rhythm in our example is slow.

But what if a dreamed rhythm *is* reported by the dreamer after awakening? If we accept that the unexpressed dreamed rhythm is inside one's head, we have no grounds to deny the same brain-bound status to the dreamed rhythm when it is expressed (although perhaps we may now introduce some *additional* externalist interpretation of the experience). For the exact same brain-bound phenomenon that justified the internalist interpretation in the unexpressed case would be also present in the expressed case—the only difference being that now the brain-bound rhythm has a causal influence on the subject's motor cortex, resulting in physical behavior that counts as a report. So, in fact there is now an additional feature of the rhythmical experience that can be attributed to the neural processes: a specific causal effect. Put simply, the only phenomenon in the entire physical universe that has both the temporal profile of the experienced rhythm (including the fact of *when* the experience occurs) and a proper causal connection to the reporting behavior are certain neural processes in the subject's head.

Let us consider next a dream about a musical melody. For the sake of simplicity, let us consider only two basic aspects of the dreamed melody: the

rhythm and the perceived pitches of the individual sounds. Let us assume also (again, for the sake of simplicity) that the rhythm of the melody is slow enough to allow one-to-one mapping between the experienced rhythm and the corresponding rhythm in the brain. Regarding pitches, there is evidence that different experienced frequencies are represented in primary auditory cortex A1 in an ordered manner, so that, roughly put, the nearby pitches are encoded by the nearby cortical areas (Da Costa et al. 2011; Oh et al. 2013). Furthermore, there is also evidence that the same frequency-to-place mapping (or tonotopy) occurs both during veridical auditory experiences and during imagined (or dreamed) auditory experiences (Oh et al. 2013). Admittedly, if we monitor the entire A1, then the tonotopical frequency-to-place mapping is far from accurate, but for the present purposes we may consider some sub-region of A1 where the mapping is accurate.

What the above considerations mean is that every musical melody has a unique two-dimensional structure and that the same two-dimensional structure (at least if monitored with a sufficiently coarse resolution) can be found in the brain of the subject who is experiencing the melody. Admittedly, the discovered frequency-to-place mapping does not explain why we experience melodies the way we do, or why we have auditory phenomenal experiences in the first place, but that does not change the fact that in the dreamed and unexpressed cases the subject's brain is the *only* place in the entire physical universe where one can find (systematically) the two-dimensional structures of the experienced melodies. And again, if such an experience is reported, then the corresponding neural processes are the only physical phenomena that have both the two-dimensional structure of the experienced melody and a proper causal connection to the reporting behavior.

It is significant that if the experienced melody belongs to a commonly known song, then the easiest way to describe its two-dimensional structure is by referring to the song. Moreover, in some real-life situations it may be the only way available. For example, if singing or whistling (or using a musical instrument) is not an option, and if the experiencing subject has no musical notation skills, then failing to recall the name of the song might result in failing to report the content of the experience. But this does not mean that the specific phenomenal aspect that is defined by its two-dimensional structure depends on the existence of the song or the corresponding concept (for the structure

could emerge in a dream even if the song did not exist). In more general terms, as long as we are concerned with purely structural properties of our phenomenal experiences, those properties can be reported and described by using concepts that refer to familiar and public phenomena with the same or roughly similar structures, but it does not follow that those public phenomena (or the corresponding concepts) are among the constituents of the (structural) phenomenal properties in question.

The simple examples considered above do not even begin to explain what the nature of phenomenal consciousness is or why it exists. Nevertheless, they prepare the ground for a hypothesis that might—a hypothesis according to which our phenomenal experiences are *fully structural* phenomena that can be found in our neural processes.

3. Toward a positive theory of phenomenal consciousness

As we saw above, it is quite safe to claim that at least some very simple and coarse-grained structural properties of our phenomenal experiences can be found in our neural processes. The suggestion that *all* structural properties of our phenomenal experiences can be found in our neural processes is, admittedly, much more ambitious, although it is defended by many prominent philosophers and scientists (e.g., Chalmers 1995, 2003; Velmans 2009; Revonsuo 2006; Tononi and Koch 2005). In order to grasp the basic idea, let us consider a speculative proposal by Antti Revonsuo (2006), who predicts that the entire structure of our phenomenal consciousness will be found in the brain once we learn to monitor the *proper level of organization* of our neural processes. According to Revonsuo, as long as we are not capable of monitoring the right level, we would find merely explanatorily insufficient neural correlates of consciousness.

In the light of the earlier examples, Revonsuo's idea can be interpreted as follows: some structural properties of our phenomenal experiences are to be found in a relatively *low level of organization* of our neural processes (e.g., the purely temporal structure of an experienced rhythm or the two-dimensional

structure of an experienced melody). This means that those properties are the structural properties of some relatively simple and easily detectable neural events. However, some other structural properties of our phenomenal experiences (in fact the majority of them) reside at much higher levels, which means that those are the structures of certain patterns of patterns of patterns, etc. of simple neural events. In order to find those higher-order structural patterns in the brain, we need powerful pattern recognition algorithms and advanced brain-imaging technology. Then, equipped with the right kinds of tools, we would eventually find a complex higher-order pattern (a pattern of patterns of patterns, etc. of simple neural events) that has a structure of the corresponding phenomenal experience.

The above proposal is speculative, but as long as it concerns only the structural properties of our phenomenal experiences, it is ultimately an empirical hypothesis that cannot be refuted by purely philosophical arguments (although it poses some general philosophical questions that are addressed further below). However, the claim that phenomenal consciousness is a fully structural phenomenon is even more radical and leads to philosophical problems that cannot be addressed by mere empirical speculations.

According to a popular view, the greatest challenge in explaining phenomenal consciousness is to explain its qualitative features—the painfulness of pain, the redness of red, the experienced scent of a rose, etc. (e.g., Chalmers 1995, 2003). Those qualitative features, the so-called *qualia*, are often claimed to be primitive and unanalyzable. What is significant in the present context is that qualia *appear* to be nonstructural: an experienced blueness is what it is, so to speak, and cannot be readily described or analyzed in structural terms like an experienced rhythm or melody.

So, at least at the outset, it seems that every individual phenomenal experience has two kinds of properties: structural and nonstructural. And while it might be legitimate to hypothesize that all structural properties of every individual conscious experience can be found in our neural processes, it seems to make no sense to suggest that the irreducibly qualitative and nonstructural "raw feels," such as an experienced blueness or the scent of a rose, could one day be found in the brain.

The apparently atomic and nonstructural nature of qualia is equally problematic for internalists and externalists. For at least according to the currently predominant physicalist conception of reality, there are simply no such things, either inside the brain or in the external world, which would be *irreducibly* qualitative in the way qualia appear to be. Therefore, it is not surprising that both internalists and externalists have made significant efforts to deny the nonstructural nature of qualia. Neuroscientifically driven internalists tend to focus on the similarity and difference relations among qualia (e.g., Churchland 1986; O'Brien and Opie 1999; Edelman 2003; Pestana 2005; Tononi and Koch 2015). Obviously, the red quale is more similar to the orange quale than to the blue quale, while at the same time the red quale is more similar to the blue quale than to the auditory quale of the sound of a trumpet. Therefore, it has been suggested that the specific quality of an individual quale is defined by its position in a complex multidimensional space that is structured along the similarity and difference relations among qualia. It is also assumed that the structure of such qualia space can be found at a certain level of neural processing.

On the other hand, a relatively recent and increasingly popular externalist approach, the so-called sensorimotor theory, aims at identifying qualia with the qualities of an organism's interactions with the environment. Although the paradigmatic examples of such a proposed identity focus on tactile qualia, such as the quale of softness (the experienced softness of a sponge is arguably a *quality* of one's interaction with the sponge while squishing it), the approach has also been applied to visual color qualia (e.g., O'Regan 2011). Put very simply, different colored surfaces have different optical responses to changes in lighting conditions. So, when interacting with three-dimensional colored objects, the visual experiences change differently depending on the color of the experienced object—each hue having its unique dynamic profile. And it has been suggested that those specific dynamic profiles are important constituents of the specific qualitative feels we associate with different experienced colors (ibid.).

It may very well be the case that there is some truth in both of the above suggestions. At least it is not obvious that the similarity relations and the dynamic profiles do not contribute anything to the experienced qualia. However, the problem with the above and other similar explanatory attempts

is that they leave a significant unexplained residue. It is one thing to accept that the location within a qualia space and a unique dynamic profile are among the constituents of the blue quale, and quite another to argue that there is nothing more to the experienced quality of blueness than those simple structural features.

So, how could one proceed from here? The suggestion proposed and defended in this dissertation is that qualia are indeed *fully* structural, but that the simple structural features acknowledged above form (at best) only a tiny fraction of their structures. According to the suggestion, the major part of the structure of any individual quale is formed by a vast network of interrelated unconscious associations. The precise nature of those associations is best understood in neuroscientific terms (specified shortly below), but for an initial intuitive grasp, we may think of them as simple psychological associations (e.g., redness is associated with blood, tomatoes, sunsets, etc.).

The above idea was originally put forward by Francis Crick and Christof Koch (Crick and Koch 1998, 2003; see also Koch 2004), but its philosophical significance has remained largely unrecognized—perhaps even by the authors themselves. For although Koch's later work with Giulio Tononi (e.g., Tononi and Koch,2015) develops Crick and Koch's original theory in many respects further, the idea of unconscious associations as constituents of qualia is no longer emphasized (the resulting explanatory loss is considered further below).

Since the theory of Crick and Koch is neurobiological, it offers a clear (although somewhat speculative) interpretation of qualia in neural terms. In a nutshell, according to their hypothesis, the unconscious associations that define the specific qualitative character of an individual quale are realized by direct (from neuron to neuron) axonal connections between small groups of neurons, the so-called *essential nodes*, so that each essential node is responsible for the detection of a feature or an object which the corresponding association concerns (e.g., Koch 2004: 34–35).

The notion of an essential node is related to the more familiar notion of a neural correlate in the following straightforward manner: the neural correlate of, say, redness is an increased activity in the essential node for redness. However, it should be noted that the notion of a neural correlate is more

flexible than the notion of an essential node. For example, we may talk about neural correlates of complex experiences, whereas an essential node is always related to a simple experiential aspect. That said, there are also essential nodes for some complex objects (such as tomatoes or grandmothers), but the neural correlates of the real-life experiences of such complex objects would usually involve several essential nodes. For example, the neural correlate of an experience of a tomato would probably involve essential nodes for tomatoes, redness, roundness, etc.

In any case, in the light of Crick and Koch's hypothesis, an individual experience of redness is a complex structural phenomenon whose structure may include the following components: it is more similar to orange than it is to blue (as most internalists emphasize), it is associated with tomatoes, strawberries, sunsets, blood, my first car, an order to stop, a need to be careful, poison, the door of my house, roses, the day of my graduation, etc., etc., etc. The associations differ in strength, and the great majority of them are unconscious, which means that an experiencing subject has no conscious access to the defining associational structure of her qualia. Nevertheless, each association contributes something to the specific character of the quale it participates in, and the totality of all the associations fully defines the quale's qualitative character. Put simply, a quale consists of a vast amount of tacit information made instantly available (although introspectively unanalyzable) to the experiencing subject (e.g., Koch 2004: 233–235).

As we saw above, the defining structure of a particular quale can be described in terms of associations with external and public phenomena (strawberries, tomatoes, the day of graduation, etc.). However, each of the external components involved is relevant only as far as it *means* something to the experiencing subject. For example, an association with tomatoes contributes to the character of *my* red quale only as far as tomatoes mean something *to me*. And the specific meaning of tomatoes *to me* is defined by all the associations tomatoes have *for me* (they can be eaten, they grow in my grandmother's greenhouse, they can be used in salads, they are usually red, I ate one on my graduation day, etc.). It follows that the overall architecture of the network of qualia is holistic: the red quale is defined by multiple associations, among which is an association with tomatoes, and the meaning of tomatoes to me is defined by multiple associations, among which is an association with redness.

According to Crick and Koch's hypothesis, that holistic architecture can be found in our brains. Neurobiologically speaking, the structure of a red quale would be centered around the essential node for redness, which projects to all the essential nodes for the associated phenomena (the essential nodes for tomatoes, strawberries, blood, etc.), and each of those associated essential nodes would project, in turn, to all the essential nodes for the phenomena they are associated with. And so on.

Initially, one might find the overly holistic architecture problematic and try to avoid it by adopting a similar but more externalist stance (e.g., Beaton 2013). For example, if the red quale could be defined by associations with some public and familiar phenomena (e.g., with *real* tomatoes and *real* strawberries), then it might feel that the corresponding theory of qualia could be, in a sense, anchored to those public and familiar phenomena. However, as we saw above, in the present context it does not matter what the associated public phenomena are *in themselves* (e.g., what their true physical nature is) or even what their established meaning is in our linguistic community (it is not part of the tomatoes' established meaning that I ate one on my graduation day), but only what they mean to the experiencing subject. And their meanings to the subject are to be analyzed in terms of some further associations with some further phenomena that are, again, relevant only as far as they mean something to the subject, and so on.

Admittedly, in the context of philosophy of language, the above use of the notion of meaning might strike as odd and idiosyncratic. For example, one might protest, in the spirit of Putnam, that meanings are simply not in the head, and that an account of qualia that leans so heavily on the concept of meaning is already an externalist account; or alternatively, in the spirit of Wittgenstein, that it makes no sense to talk about the meaning of a tomato to *me and me alone*, for meanings are never private and can exist only within some linguistic society.

However, as the above paragraphs hopefully clarify, the notion of meaning should be understood in the present context in terms of associations that are personal in the sense that they vary from person to person. So, perhaps in the context of philosophy of language it would be more appropriate to talk about *ideas* or *impressions* instead of meanings, but Crick and Koch talk about meanings (Crick and Koch 1998, 2003; Koch 2004: 242–244), and the way they

do it is compatible with an everyday use of the term: the Second World War *means* very different things to different people, and so does the national anthem of France. Similarly, tomatoes mean something different to me than they mean to you in the sense that they are associated with different things for me than they are for you. Such meanings are not necessarily private in the problematic Wittgensteinian sense, but they are *personal*. And such personal meanings (or whatever we decide to call them) are very real phenomena: different people really do associate different things with the same idea or concept.

What is important for the present purposes is that the personal and specific meanings that define individual qualia arise and remain within a holistic system and cannot be anchored, ontologically speaking, to anything external. Once again, if the red quale is defined by associations with red objects, those objects are relevant only as far as they mean something to me, and those meanings are personal in the sense that they are defined by some further associations for me, and so on. The problematic absolute privacy is avoided due to the fact that each personal associational network is defined by its unique structure that can be described in principle (more or less accurately, more or less formally, and in more or less detail).

Since the holistic architecture outlined above is a fully structural phenomenon, its description could be given, at least in principle, in terms of nodes that are connected to each other with varying strengths. And in our brains, the same architecture would be realized by essential nodes connected to each other with axonal bundles. According to this view, each of us would have a personal qualia space with a unique structure that changes over time (with new associations forming, some associations strengthening, others weakening, and some fading away). As seen above, it is possible to describe such structures in an informal manner, for example, by talking about associations with different public phenomena (tomatoes, strawberries, etc.) and specifying the personal meanings of those phenomena in terms of some further associations, and so on. But since the qualia spaces (and the individual qualia that arise within them) are fully structural phenomena, it would be possible, at least in principle, to describe them also formally—without mentioning any of the external objects. The situation is analogous to the earlier considered case of an experienced melody. Although the melody is defined by a specific two-

dimensional structure that can be described formally, often the easiest and most natural way to describe it is by referring to the familiar song the melody belongs to.

To summarize, one of the most notable virtues of Crick and Koch's hypothesis is that it offers a way to analyze qualia in fully structural terms. Moreover, the proposed structures of qualia are such that they can be understood in easily graspable phenomenal (or psychological) terms, such as unconscious associations, as well as in well-defined neurobiological terms, such as neural correlates and essential nodes. Put briefly, the hypothesis allows us to make sense, both intuitively and scientifically, of the claim that qualia are fully structural phenomena that reside literally inside our heads.

As mentioned earlier, the more recent work of Tononi and Koch, the so-called *integrated information theory* (IIT), inherits many of its central features from the earlier theory of Crick and Koch. For example, according to Tononi and Koch (2015), qualia are *maximally irreducible conceptual structures*—shapes in a "fantastically high-dimensional cause–effect space specified by a complex of neurons in a particular state." The above view (and IIT in general) is fully compatible with the idea that such shapes can be found, at least in the case of conscious human beings, within the networks of essential nodes in the brains. However, Tononi and Koch do not suggest that the "shapes of qualia" should (or could) be understood in terms of unconscious associations—or in any other easily graspable phenomenal terms. Instead, they focus on the idea that the similarity and difference relations among qualia correspond to the measurable distances between qualia-shapes in a qualia-space. But even if true, the above idea does not help us to bridge the famous explanatory gap between physical phenomena (e.g., neural processes) and qualia, for one could still ask: given all the similarities and differences between qualia and the corresponding distances between qualia-shapes, why does this particular shape give rise to that particular quale?

Initially it might seem that the same problem also arises for Crick and Koch's theory. For although the theory *claims* that the specific qualitative character of a blue quale is *fully determined* by its specific associational structure, to the experiencing subject it would remain unclear why the same associational structure could not as easily determine the specific qualitative character of a green quale instead. However, I argue that Crick and Koch's theory leads to

an empirical prediction, according to which the defining structures of qualia can be revealed to the experiencing subjects with the help of proper technology.

Let us begin by acknowledging that in some cases the defining structures of qualia can be revealed to experiencing subjects *without* brain-stimulating technology. For example, the quale of a low guitar sound is experienced initially as primitive and atomic, but reveals its structural nature once one learns to distinguish the individual overtones (Dennett 1991: 49–50). When that happens, the subject recognizes the quale as the same one she experienced before, only now she is aware that its specific overtone structure (partly) *determines* its characteristic guitarishness. In other words, after learning to distinguish the individual overtones, the typical subject does not feel as if the unitary nonstructural sound has been replaced by a new ensemble of sounds. Rather, she feels that she has become aware of the sound's internal (overtone) structure that was there all along although she failed to notice it before.

What is significant about the above example is the nature of the process it reveals: the process starts with an apparently atomic and irreducibly qualitative quale but leads to a genuine phenomenological recognition of its "internal" structure, which determines its specific guitarish character and distinguishes it from trumpetish, flutish, violinish, etc. qualia. A similar process may occur, perhaps even more effortlessly and naturally, with smells, tastes or moods: an ineffable yet somehow specific sense of anxiety may suddenly reveal itself as a worry about next week's exam (or a mixture of worries about different things), while an apparently primitive and robust gustatory difference between two brands of ketchup may suddenly be recognized as a difference in their sugar to vinegar ratio. Again, what is important about the above examples is the subject's understanding that the revealed structure determines (at least partly) the resulting quale—and therefore it is no longer possible to imagine that the same structure could have resulted in entirely different quale instead. And presumably, it would be equally impossible to imagine that the same quale could have resulted from some entirely different structure.

As the above examples suggest, the felt character of a quale may depend on various different types of associations, some corresponding to the intuitive notion of association better than others. For example, the guitarish quale

would be partly determined by its overtone structure (the "associations" to its individual overtones), partly by some "proper" associations to the subject's previous encounters with guitar music, and partly by other types of associations that remain to be discovered.

We noted earlier that phenomenal experiences seem to have two types of properties: structural and nonstructural. We noted also that while it is legitimate to hypothesize that the structural properties can be found in the brain, the nonstructural ones pose a problem. But the above examples show that at least some of the phenomenal properties that appear to be nonstructural at the outset are in fact structural in the sense that they have internal and phenomenally accessible structures that determine (at least partly) their specific qualitative character. In the light of Crick and Koch's theory, it can be predicted that such a "structure-exposing" procedure can be performed on whichever individual quale we choose to pick.

In order to see that, let us briefly consider what goes on inside one's brain, according to Crick and Koch's theory, when one is experiencing red (a more detailed account can be found in the article succeeding this introductory essay): The activity in the essential node for redness is significantly increased. Since the essential node for redness projects directly to the planning modules of the brain, the subject can now easily adjust her behavior according to the fact that something red appears to be nearby (e.g., she could stop at the red light or report the fact that she is experiencing red; see Koch 2004: 242–245). At the same time, the essential node for redness projects also to all the essential nodes associated with redness (essential nodes for tomatoes, strawberries, blood, etc.). The activity in all the associated nodes would increase slightly, but in most of them not enough to have a significant impact on the planning modules (where all of the associated nodes also project), and thereby the subject would fail to report most of the associations that jointly compose her red quale. The combined impact from all the associated essential nodes on the planning modules would still be significant, and it could affect the subject's behavior in several ways: for example, by shaping her attitude toward the experienced object or by causing her to report that the experienced redness has a particular qualitative *feel* that is ineffable yet somehow specific.

To put it very bluntly, the great majority of the individual associations that collectively compose the red quale are unconscious simply because they are

too weak. But according to the theory, each of the unconscious associations could become conscious in principle—if only the activity in the corresponding essential node increased enough. Therefore, if we were able to locate the essential nodes that are in the key position of determining the characteristic feel of redness (and distinguishing it from the characteristic feels of blueness and greenness), then we could increase their activity and make the corresponding associations conscious. According to the prediction, the result would be similar to the case of recognizing the structural nature of experienced guitarishness.

According to the emerging overall picture, every individual phenomenal experience is characterized by several qualitative properties that appear to be nonstructural at the outset, but none of those apparently nonstructural properties are *irreducibly* nonstructural; their distinguishing qualitative character is determined by the internal structure that can be revealed, at least in principle, to the experiencing subject. During such a revealing process, the revealed individual associations that compose the quale involve qualitative elements themselves, but each of those qualitative elements could be exposed as structural in the course of some further revealing process, and so on. The overall process would reveal a complex network with the holistic architecture outlined earlier: each individual quale can be seen as a "personal meaning" of an experienced aspect (e.g., an experienced blueness, painfulness, tomatoness, etc.), a meaning that is defined by a vast network of unconscious associations whose personal meanings are defined by some further associations, and so on.

To sum up, in the light of the above hypothesis, every phenomenal experience is a fully structural phenomenon whose structure could be given a formal description, at least in principle. It follows that it is legitimate to hypothesize that the structures in question can be found at a certain level of organization of our neural processes. Some of the structural features are easily recognized as such by the experiencing subject herself (e.g., the two-dimensional structure of a melody) and can be described to the others by "recreating" the structures within the public sphere (e.g., by singing or whistling the melody) or by referring to some public phenomena with similar structures (e.g., a famous song). The defining structures of qualia, however, would initially be hidden and unconscious, but can be made (partly) conscious by deliberate introspective efforts (e.g., in the case of the taste of ketchup or the feeling of

anxiety), by specific exercises (e.g., guitarishness, trumpetishness), or in the toughest cases, which concern qualia that remain impenetrable to all technologically unaided introspective methods (e.g., redness, blueness), by a specific kind of direct brain stimulation.

The above hypothesis has great explanatory potential. If it can be established that phenomenal consciousness is a fully structural phenomenon, then finding its structure in the natural world and revealing its details, as well as its relations to other phenomena, becomes a relatively straightforward empirical matter. The biggest obstacle to conceiving phenomenal consciousness as a fully structural phenomenon is the fact that it has some properties that *appear* to be nonstructural to the experiencing subject—and no mere speculation regarding their structural nature, no matter how scientifically plausible, could break that intimate and powerful impression. That is why Crick and Koch's theory is so significant: it leads to an empirical prediction, according to which the impression of non-structurality can be scattered regarding whichever individual quale we choose to pick. However, according to the view, one could never experience one's immediate phenomenal experience as a *fully* structural phenomenon. The revealed structures would always contain some qualitative and nonstructural components, but none of those qualitative components would no longer appear as *irreducibly* nonstructural, for the experiencing subject would understand that the structure-revealing process could also be applied to those.

4. Are structures real enough?

The hypothesis outlined above puts a heavy explanatory burden on the notion of structure. The examples regarding the structures of experienced rhythm and melody might be intuitive and clear enough; in those cases it is (hopefully) easy to understand what is meant by the structure and in what particular sense it can be said that certain structural properties of phenomenal experiences can be found in the brain. Perhaps what facilitates the understanding in those cases is that we are not talking about *just* structures, but certain *well-defined kinds* of structures: *temporal* structures and sorts of rough *proximity* structures (spatial

proximity in some regions of A1 corresponds roughly to pitch proximity in phenomenology). However, when moving to the claim that phenomenal consciousness is a fully structural phenomenon, we are entering philosophically deeper waters where new kinds of questions and objections may arise.

The view defended in this dissertation presupposes that it is legitimate to hypothesize that the structures of our phenomenal experiences reside in our brains in some *perfectly objective sense*. What the view ultimately suggests is that it is an objective factual matter that the structure of an individual phenomenal experience is *identical* with some structure in our neural processes. However, it has been argued that the structural identity (or structural isomorphism) and structural similarity between different kinds of phenomena depend on a method of projection, and that two phenomena that are structurally similar (or identical) under one method of projection might be entirely dissimilar under another (e.g., O'Regan 2011; Lagerspetz 2002). Simply put, it can be argued that structures are, at least to some extent, in the eye of the beholder.

In this dissertation, the above kind of criticism is addressed from the standpoint of *structural realism*. According to the theory of Ladyman and Ross (Ladyman et al. 2007), all objectively real phenomena (objects, laws, processes, events, etc.) discovered and recognized by the empirical sciences can be analyzed in fully structural terms. In practice, those phenomena are discovered as certain structural patterns in some data. However, the above does not imply that the discovered structures are any less real (or ontologically different) than the data that "contains" them; although every theory treats some of its basic data elements as a sort of nonstructural given, those data elements can always be analyzed in fully structural terms by some other theory (Ladyman and Ross 2013). In other words, one theory's structural discovery is another theory's nonstructural data.

Therefore, according to the above version of structural realism, although the world itself (at least as far as it can be known by science) is fully structural, no theory is fully structural in the sense that every theory contains some basic data elements whose nature is not analyzed in structural terms within that theory. For example, if a neurobiological theory of consciousness is ontologically committed to the discovered structure of phenomenal

consciousness, then it is equally committed to some simple neural events (e.g., action potentials, synaptic effects, firing rates, etc.) that serve as its basic data. However, if phenomenal consciousness is indeed a fully structural phenomenon (as argued in this dissertation), then we are primarily interested in finding and studying its structure. The question about the detailed nature of the individual data elements that realize that structure might be interesting in its own right, but it does not have to be considered within the theory of consciousness proper.

Assuming that individual data elements are treated in the context of the theory they belong to as sorts of nonstructural givens, the earlier considered problem takes the following form: given certain data (e.g., a set of action potentials, synaptic effects, etc.), can we say that certain particular (and *only* those particular) structures are in that data in an objective sense (i.e., independently of any actual investigative and analytical methods one chooses to apply to the data)? According to Ladyman and Ross, we can. According to their theory, there are several objective criteria (defined in information-theoretical terms) that a structural pattern must satisfy in order to count as a genuinely real object of science—the so-called *real pattern*. And if a pattern satisfies those criteria, then it is present in the data regardless if it is ever discovered or not; in other words, it is real in an objective sense.

The most basic of these criteria states that every real pattern must compress information in the data. In other words, a real pattern must encode the information in the data more efficiently (in the information-theoretical sense) than the "bit-map" encoding of that data (Ladyman et al. 2007: 226). Intuitively, this requirement corresponds to the fact that scientists do not accept every arbitrary pattern in data as a genuine scientific discovery. In practice, "discovered" patterns that do not compress information in the data would be considered nothing more than artifacts of one's imagination.

According to another criterion, a genuinely real pattern must be such that there are no other patterns that would compress the same information in the same data more efficiently (Ladyman et al. 2007: 231–233). Intuitively, that requirement corresponds to the scientific attitude according to which there can be only one correct theory of any given phenomenon. For example, the idea that Einstein's theory of gravity is more accurate than Newton's theory of gravity presupposes that there can be only one correct theory of gravity and

that Einstein's theory is a more accurate approximation of that final theory. Correspondingly, scientists *do not* believe that Newton's theory describes one type of gravity and Einstein's theory another. It is assumed that there is only one kind of gravity that science seeks to explain and describe. In the light of structural realism, there is only one structure that captures the "true nature" of gravity—and that structure exists in the world in an objective sense, regardless if it is ever discovered or not.

Obviously, gravity is not the only real pattern in the world, but if it is genuinely real, then it is *indispensable* in the sense that replacing or removing it from a theory would come with a loss of compressibility in some "real" data (i.e., data that consists of real patterns). According to Ladyman et al. (ibid.), all real patterns are indispensable in that sense (the claim is implicit in the requirement that there must be no other patterns that would compress the same information in the same data more efficiently).

Considering the complexity of the human brain, it is highly likely that the neural processes of a conscious person contain several real patterns that are indispensable in the above sense. According to the hypothesis defended in this dissertation, some of those indispensable structural patterns are the structures of our phenomenal experiences. One might now wonder why those particular structures are phenomenally conscious and not others. However, in light of the hypothesis defended in this dissertation, such a question would be misguided, for it would presuppose that the structure of our phenomenal consciousness *is* phenomenally conscious. But according to the hypothesis, consciousness is a fully structural phenomenon (i.e., having no properties over and above its structural properties). If an individual conscious experience *is* a certain structure, then once we have found that structure, we have found the conscious experience itself—with *all* its properties. Thus, it would make no sense to ask why that particular structure is phenomenally conscious. The temptation to ask such a question originates from a strong and persistent intuition, according to which the structure of our phenomenal experiences is somehow "filled" with irreducibly qualitative and nonstructural qualia. But if those qualia can be analyzed in fully structural terms, then the "filling" becomes part of the structure.

To sum up, the version of structural realism defended by Ladyman et al. (2007) supports the view that certain structures exist in the world in an objective

sense. This view does not deny that many structures are artificial constructs that may (or may not) have an instrumental value, but those structures are not, according to the view, indispensable, and therefore they are not real patterns in an objective sense. Needless to say, the hypothesis defended in this dissertation assumes that phenomenal experiences are real patterns and not merely useful patterns in an instrumentalist sense.

Before proceeding, it is worth acknowledging that the above information-theoretical notion of a real pattern is fully compatible with the naïve notion of structure used in the examples of experiencing rhythm and melody. For example, if an experienced flashing of a red light has a particular rhythm (i.e., a temporal structure), then the corresponding neural rhythm would be a real pattern in data consisting of individual neuronal firings in a certain cortical region. When a subject is seeing red, the *average frequency* of individual firings in that cortical region is significantly higher than when she is not seeing red. Periods of high frequency alternate with periods of low frequency, and the information concerning those alternations can be expressed more efficiently by describing the rhythm rather than by giving an account of all individual firings (which would count as a "bit-map" description of the same information).

There are other ways in which the framework of structural realism may contribute to the debate about phenomenal consciousness. For example, from the point of view of structural realism, it is perfectly legitimate to hypothesize that many structural properties of *veridical* experiences can be found simultaneously both within the brain and in the environment (or in an organism's interactions with the environment). If so, when describing our *veridical* experiences we would indeed describe some objectively real external structures that are explanatorily significant for different versions of consciousness externalism. But since we would describe *only structures*, we would also describe at the same time the identical structures in our neural processes. And in the case of dreams and hallucinations, we would describe certain structures that can be found *only* in our brains, but we would still describe them (both to others and to ourselves) in terms of external and public phenomena.

The above ideas allow us to accept two popular externalist theses while remaining committed to the internalist view. It might be true that veridical

experiences are in some more or less literal sense out there in the environment (e.g., in a form of experienced objects or in a form of an organism's interactions with the environment), and it might also be true that we are bound to understand our non-veridical experiences in terms of external and public phenomena. Nevertheless, we may say that our experiences are inside our brains in the sense that only there, according to the hypothesis, can we find the *structures* of *all* our experiences—both veridical and non-veridical, both reported and unreported.

One of the initial motivations behind structural realism was the desire to understand the nature of radical theory changes. The history of science is rich with instances where a predictively and explanatorily successful theory has been replaced by a new theory with entirely different ontological content. Considering some famous examples, the notions of *ether*, *phlogiston* and *gravitational force* have been replaced by the notions of *electromagnetic field*, *oxidation* and *curvature of spacetime*. According to structural realism, neither the replaced nor the replacing notions refer to anything real, but are merely heuristic devices that help scientists to think and talk about the structural content of their theories (see Worrall 1989; Ladyman 2014).

According to structural realism, the so-called radical theory changes are not as radical as they might seem, for there is clear structural continuity between the replaced theories and the ones that have replaced them (see Worrall 1989; Ladyman 2014). The only thing that changes radically is the quasi-ontological content, which may have *great heuristic value* but does not correspond to anything real.

In light of the above ideas, one of the primary goals of consciousness research would be to find the structures of individual phenomenal experiences in some well-defined data. At the very least, the theory of consciousness should offer an empirical prediction in what kind of data the structures of individual experiences can be found. The question of which vocabulary or what linguistic mechanisms can or should be used to describe those structures is a secondary matter that belongs to the domain of heuristics. As argued in this dissertation, the way most externalist accounts approach *non-veridical* experiences confuses the ontologically significant structural issues with the matters of heuristics. More specifically, externalists offer no prediction of where (in what kind of data) one would find the highly specific structures of individual unreported

dreams if not in the brains of dreaming subjects. The fact that we are bound to describe the content of our dreams in terms of public and external phenomena (at least in an informal context) may have several interesting consequences, but it does not threaten the hypothesis that the structures thus described are certain objectively real structures in our neural processes.

As already noted, the main reason to prefer the brain-bound internalist hypothesis (regarding consciousness) over the externalist view is the existence of certain non-veridical experiences whose specific content does not correspond to anything in the external world (e.g., Tononi and Koch 2008; Revonsuo 2006, 2015; O'Regan and Block 2012; Searle 2000). As argued above, most externalist accounts approach such non-veridical experiences from the perspective of philosophy of language—by appealing, one way or another, to the fact that we think and talk about our non-veridical experiences in terms of public and external phenomena. However, that is not the only externalist strategy regarding the non-veridical experiences. For example, it has been suggested that dreams might not exist, and that our impression that they do is based on false memories generated upon awakening (Dennett 1976). Alternatively, it has been argued that although some non-veridical experiences might require an internalist explanation, the veridical experiences must nevertheless be explained within the externalist framework (e.g., Hurley 2010). But perhaps most often it is simply claimed that dreams are phenomenally much poorer (i.e., less vivid, less detailed, less coherent, etc.) than veridical experiences, and therefore have little or no relevance when it comes to explaining the latter (e.g., Noë 2009; O'Regan 2011; Hutto and Myin 2013).

The above ideas do not depend on the arguments from philosophy of language. In light of structural realism, we can interpret them as follows: Dennett's suggestion would mean that the structures of our dreams can be found *nowhere*—at least during the time we are asleep. Hurley's suggestion would mean that it might be the case that the structures of our dreams can be found in our brains, but the structures of our veridical experiences cannot. The appeal to the alleged phenomenal poverty of dreams would support Hurley's idea by suggesting that although the simple and primitive structures of our dreams might be found in our brains, the rich and complex structures of our veridical experiences cannot.

Although these ideas do not in themselves appeal to philosophy of language, they are almost always presented conjointly with the latter. However, once separated from linguistic issues, as structural realism encourages us to do, the above suggestions reduce to relatively simple and straightforward empirical matters. And as argued in this dissertation, the results of neuroscience and empirical dream research provide us with plenty of reasons to doubt their empirical plausibility.

In brief, Dennett's idea has been falsified by multiple experiments on lucid dreaming (e.g., LaBerge et al., 1981; see also Dresler et al., 2012), as well as by the so-called "mind reading" experiments (Horikawa et al. 2013; Siclari et al. 2017; see also Sebastian 2014). Hurley's view depends on the idea that there are specific kinds of major differences in neural processing during dreams and veridical experiences. And as argued in this dissertation, there are no empirical signs of such specific kinds of neural differences. Finally, the popular claim that all dreams are phenomenally poorer than full-blown veridical experiences has been refuted by numerous results of empirical dream research (e.g., Revonsuo and Salmivalli 1995; Domhoff 2007; LaBerge and DeGracia 2000; Nir and Tononi 2010; Hobson 2009; see also Revonsuo 2006: 79–84). A more detailed treatment of the issues outlined above is provided in the second article further below.

5. Summary

The view of phenomenal consciousness as a fully structural phenomenon has great explanatory potential. To begin with, the view allows us to separate the questions of *what* is the nature of some particular phenomenal experience and *where* we can find that experience from the questions of *what kind of vocabulary* can one use to describe the experience and *in what kind of terms* can one *understand* or *think about* the experience. For it is perfectly legitimate to hypothesize that even if we are bound to describe the content of our phenomenal experiences in terms of public and external phenomena, the *structures* thereby described can be found in our brains (and in some cases only in our brains).

Second, the fully structural view of phenomenal consciousness allows us to address the so-called *hard problem*: why is there *something it is like to be* conscious? If the apparently qualitative features that define the specific nature of *what it is like* to be conscious are fully structural, then it becomes possible to study and explain those features by tracking their structures in the natural world.

As argued in this dissertation, the theory of Crick and Koch (1998, 2003; see also Koch 2004) allows us to understand the structural nature of qualia in both well-defined neural terms (essential nodes and their axonal connections) and in easily graspable phenomenal terms (unconscious associations). Moreover, if Crick and Koch's theory is approximately true, then it should be possible to reveal the structural nature of whichever individual quale to the experiencing subject. The idea that such a structure-revealing process can actually occur is supported by the cases in which a similar process occurs without the help of brain-manipulating technology.

According to the hypothesis, the subject would never experience her qualia as *fully* structural, for the revealed structures would always contain some apparently nonstructural components. However, each of those qualitative elements could be "structuralized" in a similar manner as the original quale, and so the subject would realize that none of her qualia are irreducibly qualitative. The nature of the above revealing process could be captured by the slogan: "Show me an irreducibly qualitative quale and will I show you the unconscious associations that determine its qualitative nature."

The above hypothesis presupposes that some structures exist in the world in an objective sense—independently of their possible instrumental values or uses. As argued in this dissertation, one possible way to defend—as well as clarify—the above position is by appealing to a version of structural realism. According to the theory of Ladyman et al. (2007), if a structure satisfies certain criteria (definable in information-theoretical terms), then it exists in the world in an objective sense, independently of anybody's knowledge of its existence or the possible instrumental roles it might play in some particular scientific theories. As suggested in this dissertation, phenomenal experiences are that kind of objective structures residing in our neural processes.

The main themes of this dissertation are distributed in the subsequent articles as follows:

The first article focuses on the positive thesis by introducing and defending the idea that conscious experiences, including the qualia, are fully structural phenomena residing in our neural processes.

The second article focuses on the shortcomings of the currently popular sensorimotor theory by criticizing its approach toward non-veridical experiences (especially dreams). That article contains many specific arguments against different sensorimotor explanatory strategies. Those arguments are self-sufficient in the sense that they work independently of the general structuralist framework defended in this dissertation, and for that reason they receive less attention in this introductory essay. However, their role in the dissertation is important, for as mentioned above, some of them target externalist explanatory strategies that do not appeal to philosophy of language. Simply put, those externalist strategies demand special attention, for they cannot be disarmed or dismissed by separating structural issues from linguistic ones.

The third article focuses on the prospects of structural realism as a framework for analyzing the dispute between consciousness internalism and externalism. Instead of arguing for a fully structural nature of phenomenal consciousness, the article emphasizes the fact that at least some properties of our phenomenal experiences are clearly structural. Based on that, it is argued that the only phenomena where one can hope to find all the structural properties of all the phenomenal experiences (of human beings) are neural processes of the experiencing subjects. It is argued further that phenomenal experiences cannot reside in phenomena that lack some of their structural properties. In addition, it is argued that structural realism allows us to analyze the notions of structure, structural identity and structural similarity without relying on subjective methods of projection. In other words, it is argued that some structures exist in the world in an objective sense.

References

Beaton, M. 2013. Phenomenology and embodied action. *Constructivist Foundations* 8 (3): 298–313.

Chalmers, D. 1995. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2 (3): 200–219.

Chalmers, D. 2003. Consciousness and its place in nature. In *The Blackwell guide to philosophy of mind*, ed. S. Stich and F. Warfield, 247–272. Oxford: Blackwell.

Churchland, P. M. 1986. Some Reductive Strategies in Cognitive Neurobiology. *Mind* 95 (379): 279–309.

Clark, A. 2009. Spreading the joy? Why the machinery of consciousness is (Probably) still in the head. *Mind* 118: 963–993.

Crick, F., and Koch, C. 1998. Consciousness and neuroscience. *Cerebral Cortex* 8: 97–107.

Crick, F., and Koch, C. 2003. A framework for consciousness. *Nature Neuroscience* 6: 119–126.

Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., Saenz, M. 2011. Human primary auditory cortex follows the shape of Heschl's gyrus. *Journal of Neuroscience* 31 (40): 14067–14075.

Dennett, D. 1976. Are dreams experiences? *Philosophical Review* 73: 151–171.

Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.

Domhoff, G. W. 2007. Realistic simulation and bizarreness in dream content: Past findings and suggestions for future research. In *The New Science of Dreaming*, vol. 2., ed. D. Barrett and P. McNamara, 1–27. Westport, CT: Praeger.

Dresler, M., Wehrle, R., Spoormaker, V. I., Koch, S. P., Holsboer, F., Steiger, A., et al. 2012. Neural correlates of dream lucidity obtained from

contrasting lucid versus non-lucid REM sleep: A combined EEG/fMRI case study. *Sleep* 35: 1017–1020.

Dretske, F. 1995. *Naturalizing the mind*. Cambridge: MIT Press.

Edelman, G. M. 2003. Naturalizing consciousness: A theoretical framework. *Proceedings of the National Academy of Sciences of the USA* 100 (9): 5520–5524. doi: 10.1073/pnas.0931349100

Hobson, J. A. 2009. REM sleep and dreaming: Towards a theory of protoconsciousness. *Nature Reviews Neuroscience* 10: 803–813.

Horikawa, T., Tamaki, M., Miyawaki, Y., Kamitani, Y. 2013. Neural decoding of visual imagery during sleep. *Science* 340: 639–642.

Hurley, S. 2010. Varieties of externalism. In *The extended mind*, ed. R. Menary, 101–154. Aldershot: MIT.

Hutto, D., and Myin, E. 2013. *Radicalizing Enactivism: Basic Minds Without Content*. Cambridge, MA: MIT Press.

Koch, C. 2004. *The quest for consciousness: a neurobiological approach*. Englewood: Roberts and Co.

LaBerge, S., and DeGracia, D. J. 2000. Varieties of lucid dreaming experience, in *Individual Differences in Conscious Experience,* ed. R. G. Kunzendorf and B. Wallace, 269–307. Amsterdam: John Benjamins.

LaBerge, S., Nagel, L. E., Dement, W. C., Zarcone, V.P. 1981. Lucid dreaming verified by volitional communication during REM sleep. *Perceptual and Motor Skills* 52: 727–732.

Ladyman, J., and Ross, D. 2013. The world in the data. In *Scientific metaphysics*, ed. D. Ross, J. Ladyman, and H. Kincaid, 108–150. Oxford: Oxford University Press.

Ladyman, J., Ross, D., (with Spurrett, D., and Collier, J.) 2007. *Every thing must go: metaphysics naturalized*. Oxford: Oxford University Press.

Lagerspetz, O. 2002. Experience and consciousness in the shadow of Descartes. *Philosophical Psychology* 15: 5–18. doi: 10.1080/09515080120109388

Lycan, W. G. 1996. *Consciousness and Experience.* Cambridge, MA: Bradford Books / MIT Press.

Myin, E. 2016. Perception as something we do. *Journal of Consciousness Studies* 23 (5–6): 80–104.

Nir, Y., and Tononi, G. 2010. Dreaming and the brain: From phenomenology to neurophysiology. *Trends in Cognitive Sciences* 14: 88–100.

Noë, A. 2009. *Out of Our Heads*. New York: Hill & Wang.

O'Brien, G., and Opie, J. 1999. A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences* 22: 127–48

Oh, J., Kwon, J. H., Yang, P. S., Jeong, J. 2013. Auditory imagery modulates frequency-specific areas in the human auditory cortex. *Journal of Cognitive Neuroscience* 25: 175–187. doi: 10.1162/jocn_a_00280.

O'Regan, K. 2011. *Why red doesn't sound like a bell: explaining the feel of consciousness*. New York: Oxford University Press.

O'Regan, J. K., and Block, N. 2012. Discussion of J. Kevin O'Regan's "Why Red Doesn't Sound Like a Bell: Understanding the Feel of Consciousness". *Review of Philosophy and Psychology* 3: 89–108.

Pestana, M. 2005. (A Laconic Exposition of) a method by which the internal compositional features of qualitative experience can be made evident to subjective awareness. *Philosophical Psychology* 18 (6): 767–783.

Revonsuo, A. 2006. *Inner presence: consciousness as a biological phenomenon*. Cambridge: The MIT Press.

Revonsuo, A. 2015. Hard to see the problem? *Journal of Consciousness Studies* 22 (3–4): 52–67.

Revonsuo, A., and Salmivalli, C. 1995. A content analysis of bizarre elements in dreams. *Dreaming* 5 (3): 169–187.

Searle, J. 2000. Consciousness. *Annual Review of Neuroscience* 23: 557–578.

Sebastian, M. A. 2014. Dreams: An empirical way to settle the discussion between cognitive and non-cognitive theories of consciousness. *Synthese* 191: 263–285.

Siclari, F., Baird, B. Perogamvros, L., Bernardi, G., LaRocque, J. J., et al. 2017. The neural correlates of dreaming. *Nature Neuroscience 20* (6): 872–878.

Tononi, G., and Koch, C. 2008. The neural correlates of consciousness - an update. *Annals of the New York Academy of Sciences* 1124: 239–261.

Tononi, G., and Koch, C. 2015. Consciousness: here, there and everywhere? *Phil. Trans. R. Soc. B* 370**:** 20140167.

Tye, M. 1995. *Ten problems of consciousness: a representational theory of the phenomenal mind*. Cambridge: MIT Press.

Velmans, M. 2009. *Understanding Consciousness*. 2nd ed. London: Routledge

Worrall, J. 1989. Structural realism: the best of both worlds? *Dialectica* 43: 99–124.

Original Articles