The Biostatistical Model of Disease

The Role and Meaning of Values in the Construction of Health and Disease

Mikko Joonas Hartonen

University of Helsinki
Faculty of Social Sciences
Practical Philosophy

Master's thesis
March 2019

Abstract

The thesis examines various claims about value-ladenness of the medical concept of disease. The focus is on the discussion about the boundaries of normality originating in the controversies around psychiatry in the 1960s and the 1970s and currently still active. Primarily, the analysis is centered upon Christopher Boorse's Biostatistical Model (BST) which aims at setting the boundaries between health and disease. The BST is a naturalist account portraying health and disease as scientific and value-free concepts. The critics of the BST either view health and disease as essentially normative concepts or just provide claims about the BST's covert normativity ignoring the question of the correct account of the concepts.

The main question considered is whether the BST or the concept of disease it analyzes are value-laden. Secondly – but even more importantly – it is assessed what it means for a scientific concept such as disease to be value-laden and what might be the implications of such value-ladenness.

Conceptual analysis is applied throughout the thesis. The validity of different arguments for and against the value-freeness of disease and the BST is assessed.

The primary argument is that disease as defined by the BST is value-laden even though the concept may be given a description in value-free terms. The values enter on the level of choice of either the goals of biological organisms or the reference classes applied in the BST. Given different values, alternative goals or reference classes could have been chosen. Choosing alternative goals or reference classes would lead to a different concept of disease. Hence, values affect the concept of disease that results from the BST and thus disease may be considered value-laden.

The secondary argument is that while disease as defined by the BST may be considered value-laden, the value of this consideration is rather limited in the absence of further elaboration. From the observation that disease is value-laden in the way described above, it does not follow that the concept is not scientific or that the medical science and the concepts it applies are flawed. Moreover, the sort of value-ladenness in question does not necessarily imply any normative flaws in medical theory, the BST or the concept of disease.

It is concluded that if the values driving the choices medicine and the BST make are shared enough, then the value-ladenness resulting from those choices may be considered inconsequential. The problematic issues would only arise in contexts where values are diverse and competing. Determining the correct role and meaning of values in construction of medical concepts is still a work in progress.

Table of Contents

1 Introduction

1.1 The concepts of health and disease

In this thesis, the primary target of analysis is the concept of disease. The concept is of interest to various fields of study from medicine and psychiatry to different areas of social science and philosophy. My focus is very specific though*: I aim to investigate whether the concept of disease is fundamentally scientific and value-free instead of being a value-laden construct*. The value-free account of health and disease is defended by the naturalists, typically grounding the concepts in some form of biological function and dysfunction. On the opposing side of the debate are the normativists, holding that health and disease are fundamentally value-laden concepts, grounded in our norms and values.

In the context of philosophical debates about medical concepts, the focus is mostly on the concept of disease while health is typically viewed simply as *absence of disease*. By this view, if disease is biological malfunction or abnormality, it follows that a healthy person is someone whose biological systems are all in order. Thus, the concepts do not require separate analysis, as they would, if health was considered a "positive state" as it sometimes is. (Murphy 2015.) Accordingly, I focus on disease and adopt a perspective assuming health is just absence of disease. The different positive accounts of health that have been suggested (e.g. WHO 1948) are mostly ignored.

The problem of value-ladenness is more complex than it seems on the surface. This is largely because there is no shared understanding of what it exactly means for a scientific concept to be value-laden. Generally, the role and meaning of values in science depend on the context (Kincaid 2007). There are different levels of value-ladenness and different domains over which the claims of value-ladenness can be assessed (Kingma 2014). It varies quite a lot on which level and domain a writer talking about value-ladenness is focusing on. This causes misunderstandings and talking past each other.

Because of these complexities, it may be that just asserting "*disease* is value-laden" says very little. To put meaning into the claim, further elaboration is needed. For this reason, I will not *only* try to figure out whether disease should be considered value-laden or not. I will also try to shed some light into the following question: *what does it matter whether disease is value-laden in a certain suggested way*? I believe it is this question, especially,

for which coming up with a satisfying answer would currently provide the most value. Hopefully, I can contribute at least in a way of pointing to right direction in the search for this answer.

But before proceeding any further, there is one basic question to be addressed. It is about the motivation of this research project. In the first place, why should anyone be interested in what health and disease exactly are? Isn't it just a practical matter to define our concepts so that they best serve us in our interaction?

1.2 Why it matters what disease is?

If random persons on the street were asked what they think health and disease are, they would probably react by questioning the interviewer's motives: why would anyone ask such a simple question? Everyone knows what health and disease are! This reflects the fact that in common language, we seemingly share an understanding of both the meaning of these concepts and the way the terms are used in our daily interaction. But should our imaginary interviewer go on and ask how these supposedly simple and agreed concepts should be defined, the interlocutors would probably become hesitant and even if they came up with an answer, it would frequently be vague and confused[1]. Different kinds of disease processes, illnesses, injuries, disabilities and traumas with their associated pains and distresses would surely pop up into their minds. But it is unlikely that in the end our interviewer would be left with much else than a bunch of diverse and contradicting definitions, suggesting that the meanings of health and disease might not be so self-evident after all.

Philosophers are keen on rushing to help when it seems there is confusion about concepts. Conceptual analysis is one of philosophers' favorite things in the sense that it is applied so widely. Yet it is somewhat unclear what its value is (Margolis & Laurence 2014). I do not have a general stance on that question, but as conceptual analysis of health and disease is what this thesis is primarily about, I probably should have a reason to think it has value in this particular case. As it happens, I do.

Even if I am right about laypeople giving messy and contradicting definitions for health and disease and using them in confused ways, they still seem to be quite content with

[1] This may reflect the fact that people are better at using high-level concepts than at defining them (Fulford 2001, p.81).

themselves and not striving for more rigorous conceptualizations. So, one may ask, why is there a bunch of philosophers discussing and arguing about these concepts for 40 years[2]?

My precursory answer is that *it is because health and disease are of such major significance in people's lives and the concepts are thus deeply embedded in our social and institutional practices*.

People might be content with their concepts in daily communication. Without much confusion or misunderstanding, they can talk about disease as subjective experience causing stress, pain, harm or distress. The meaning of that *experience* is clear and shared, even if the concept itself is fuzzy. However, our notions of health and disease as *operationalizations* in various institutional frameworks are mostly *not* about subjective experience. Usually, the situation is exactly the opposite: our institutions apply certain, more or less rigorously defined concepts of health and disease and assume there are somewhat objective criteria for verifying whether someone is diseased or not.

Hence, the sort of operationalizations of disease our institutions are applying, affect who becomes officially considered as diseased and thus affect also people's lives in those areas the institutions in question have power over.

For example, our *health care system* is largely based on the premise that its resources should be targeted at preventing and treating diseases[3]. There are decisions made about who should be treated and who should not, and these decisions are partly grounded on disease judgements and assessments about the level of disease one is suffering from. *Health insurance companies* provide criteria one needs to fulfill if intending to get compensation for a medical problem. These criteria are largely about what diagnoses one has and how diseased physicians consider one to be. Furthermore, our *legal system* sometimes makes judgements about criminal responsibility partially based on defendant's disease status. If one has a mental disorder, one may be judged not responsible or less responsible for his criminal actions. What is considered a disease and what is not may also guide *scientific research and funding*. Disease judgements in all these areas may have significant consequences for the individuals concerned. (Boorse 2011, p.15–16; Reznek 1987, p.1.)

[2] Boorse's first paper about health and disease was published 1975. The debate is still going on.
[3] It should be noted though that health care cares not *only* about diseases (Boorse 2014, p.693).

Those mentioned above are archetypal examples of the ways disease judgements are embedded in our institutional practices. Most likely there are other ways too. What is common to all the cases presented is that in each of them a *line must be drawn*. For judgements about who is entitled to public medical treatments, insurance payments, legal excuses or research funding, our institutions have to decide where lies the boundary between health and disease. To draw the line, they need *operationalizations* of the concepts. The operationalizations must be based on *some conception of what health and disease fundamentally are*.

This is where philosophers jump in with their conceptual analysis. They work to clear contradictions, misunderstandings and fuzzy use of language. They try to refine and revise the concepts of health and disease so that their definitions would reflect the concepts´ "correct and true meanings", given certain premises that are justified separately.

The value of this philosophical conceptual work is that – in principle – it could give our institutions more clear, consistent and sound base where to build on their operationalizations of health and disease. If the operationalizations are grounded on confused base-concepts, then it is more likely that also the judgements those institutions end up making are confused and perhaps unjust, unequal or problematic in some other way. Well-defined fundamental concepts do not guarantee just judgements or fair outcomes of institutional processes, but at least they make them plausible.

1.3 Literature

This essay focuses on the conversation about health and disease which started from criticism aimed at psychiatry in the 1960s and then evolved into a deep philosophical debate about the definitions of these concepts. In the historical section, I rely mostly on the works of Hannah Decker (2013), Rachel Cooper (2015) and Greg Murray (2011). The ideas of physician John Scadding (1959; 1963; 1967) and psychiatrists Thomas Szasz (1960), Robert Kendell (1975; 2001), Samuel Guze (1978) and Robert Spitzer (1975; 1978; 2001) played a major part in setting up the table for the philosophical discussion. Szasz was a major figure in the antipsychiatry movement back in the 1960s and his works strongly motivated the conversation about health and disease in the beginning. Scadding pioneered in outlining the naturalist and statistical account of disease. Kendell, Guze and Spitzer have all contributed a lot by clarifying the meaning of *mental disorder* and the question

of why it should be understood as a subclass of *medical disorder*. Their contributions pushed forward the general understanding of the boundaries between health and disease.

In the 1970s Christopher Boorse (1975; 1976a; 1976b; 1977) worked diligently on the subject and came up with his naturalist account of disease. In the following decades it has been Boorse's original ideas – especially the Biostatistical Model (BST) – generating the liveliest debates around the concepts of health and disease (Murphy 2015). In addition to Boorse's seminal papers from the 1970s, he has given exhaustive replies to all his notable critics first in 1997 (A Rebuttal on Health) and more recently in 2014 (A Second Rebuttal on Health). Jerome Wakefield (1992; 2014; 2017) is another influential naturalist[4] whose Harmful Dysfunction Analysis (HDA) has gathered plenty of attention and whose work I sometimes refer to. The primary focus of this thesis is on Boorse's work though, as Boorse's account is more general and debated and Wakefield's analysis more narrowly targets psychiatry.

The most notable critics of Boorse's Biostatistical Model – addressing the question of value-ladenness of health concepts– are Elselijn Kingma (2010; 2013; 2016; 2017), K.W.M. Fulford (1989; 2001), Marc Ereshefsky (2009) and Scott DeVito (2000). Probably there are others, but the arguments of those mentioned are well discussed.

The section about conceptual analysis, naturalism and revisionism owes a lot to Kingma (2017), Haslanger & Saul (2006) and Murphy (2015). Kingma and Haslanger & Saul provide an enlightening view on what conceptual analysis is about and which different angles it may be approached from. Murphy has useful observations on the distinction between conservativism and revisionism about concepts and how these stances combine with naturalist and constructivist positions.

Finally, I should mention Kincaid et al. (2007) elaborating the role and meaning of values in science.

There is always some arbitrariness involved when delineating the subject of study and surely something of interest is left out. For example, there is a somewhat parallel conversation going on in the field of *disability studies* (see Vehmas & Mäkelä 2008a; 2008b) dealing with the question of whether the concepts of "impairment" and "disability"[5] are

[4] Wakefield has been called *naturalist*, *normativist* and *hybrid* depending on the perspective of the observer.
[5] A very similar distinction as Boorse's (1975) "disease" vs. "illness".

fundamentally natural or evaluative. But as the conversation seems contextually and terminologically detached from the one I am focusing on, I have decided to leave it out. Since health and disease are discussed widely within many academic disciplines and philosophical contexts, it is plausible there are other similar cases. Still, I believe the current choice of focus should be quite reasonable given the historical background and the weight of the covered authors in the literature.

1.4 Terminology

It is mostly a matter of convenience to decide which term to use when referring to a particular idea or concept. But sometimes diverse terminological customs may cause confusion and misunderstanding, especially while doing conceptual analysis. This is why I think it is appropriate to be as explicit as possible about the terminological issues in the current context.

Christopher Boorse, whose model and concept of disease are the main targets of my analysis, consistently uses *disease* in his work when he is referring to medicine's theoretical concept of disease (Boorse 1975, p.50). He deviates from this practice mostly when he is commenting the work of other people who use terms differently. Boorse has replaced *disease* with *pathological condition* in his later works to be better in line with medical writing (Boorse 2011, p.26), but these terms refer to a concept similar enough to mostly share meaning and analytical properties. Of these two, I almost exclusively use *disease* in my analysis since Boorse is practically the only one using *pathological condition* in the relevant literature.

The problem is that, overall, the usage of terms varies quite a lot. There are three terms which are frequently used when referring to "problematic physical or mental conditions" in which medicine and physicians seem interested in. These terms are *disease*, *disorder* and *illness*. The usual case is that different writers give slightly different meanings to the concepts they refer to with a term, be it any of the three mentioned. At worst, all three terms may be used as synonyms referring to the same concept (Wakefield 2017, p.69). From the analytic perspective, the problem is the inconsistency of terminological customs.

There is *some* consistency though, in the ways each of these terms is typically used in different contexts. Most commonly, *disease* refers to a class of medical conditions understood in a more naturalistic and "real" sense than conditions referred to with *illness*. Diseases are roughly those medical conditions described in textbooks and research papers by medical theorists and scientists (Boorse 1977, p.551). Boorse and his fellow naturalists usually think it is this kind of theoretical "disease" that is scientific and value-free. *Illness* more often points to a clinical or lay concept and is usually assumed to entail normative connotations. Illnesses are about people's negative experiences (Nordenfelt 2017, p.32) and judged to cause harm and deserve medical attention and perhaps treatment (Boorse 1975, p.61). Lastly, *disorder* is most typically associated with psychiatric conditions. In psychiatry it is common to speak of *mental disorders* (APA 2013). Moreover, some writers use *disorder* as a generic term referring to all deviations from health, including diseases, illnesses, injuries, traumas etc. (Wakefield 2017, p.69). Quite frequently one sees disorder applied as this kind of an umbrella concept.

Some writers have made clear and explicit distinctions between the referents of the three different terms, especially between the referents of *disease* and *illness*. For example, Boorse (1975) refers with *disease* (or later with *pathological condition*) to a theoretical scientific concept and with *illness* to a practical and normative lay concept. On the other hand, some people use *disease* and *illness* basically as synonyms without distinguishing them explicitly in any way (e.g. Kendell 1975). Szasz (1960) sometimes refers with *disease* to a very specific category of medical conditions, i.e. physical lesions and distinguishes these from "mental illnesses", which he thinks do not exist. But simultaneously, Szasz allows the possibility of "physical illnesses", which he does not explicitly distinguish from "physical diseases". Wakefield (1992; 2014; 2017) mostly operates in the context of psychiatry and speaks of *disorders*, regardless whether the question is about mental conditions or physical conditions.

Usually it is somewhat clear from the context whether an author refers to a more theoretical and perhaps value-free medical concept or to a more practical and perhaps evaluative medical concept with his preferred term. Still, many times it takes some effort to assess whether two writers are referring to the same concept with a particular term, be it *disease*, *illness* or *disorder*. This probably causes some misunderstandings.

In this essay, my focus is on Boorse's concept of theoretical disease and that is why I intend to stick with *disease* whenever the context allows. Those times the context seems to demand the use of *illness* or *disorder*, I will use the appropriate one and try to make it clear why. This happens most typically when the context is historical or when I refer to somebody using those optional terms. Hopefully, I can avoid most of the potential confusion with this terminological strategy.

1.5 Structure

I begin my analysis with a background chapter, purporting to place the discussion in context. First, I briefly take note on how the concept of disease was still largely undefined in the sixties. Then I explain the reasons why the debate about the concept gained new traction and why it was originally intertwined with issues related to psychiatry.

The third chapter is about conceptual analysis and certain distinctions relevant to my analysis. This chapter is divided in three parts: first, I assess what is the aim of the analysis targeting the concept of disease in the context of my research. Second, I present the distinction between conservativism and revisionism and explain what each position implies. Third, I explain what is meant by naturalism and normativism in the context of discussion about health concepts.

In the fourth chapter, I introduce the Biostatistical Model (BST) and have a glimpse on the previous work it builds on. I show how the model is constructed and how it defines health and disease. Then I explain its key terms. The section closes with the analysis of Boorse's distinction between *disease* and *illness*.

In the fifth chapter, I present the main types of objections the BST has faced. Then I distinguish between three different forms of criticism related to potential value-ladenness of disease. I argue against two of them, one resting on the claim that all science is value-laden and the other concentrating on the value-laden use of language.

In the sixth chapter, I engage with the most complex criticism of the BST related to value-ladenness. The aim of this section is to judge whether the BST and/or medicine are value-laden in the level of choosing goals or reference classes. Even more importantly, my goal is to assess the *meaning of value-ladenness* in the context of health concepts.

Finally, I conclude the thesis by summing up the results, discussing the factors limiting any inferences made and suggesting a way forward.

2 Background

2.1 The pressures for defining disease

For much of the modern era, there have been two rivaling conceptions of disease. On the one hand, disease has been understood as just an observable suite of symptoms with a predictable course unfolding. This notion dates back to Sydenham in the seventeenth century and Emil Kraepelin (1856–1926) applied it in psychiatry in the absence of deeper scientific knowledge of etiology or psychopathology. (Murphy 2015; Decker 2013, p.38.) On the other hand, diseases have been seen as "destructive processes in bodily organs which divert part of the substance of the individual from the actions which are natural to the species to another kind of action" (Murphy 2015). This second – seemingly naturalist – way of conceptualizing disease is still typical of medical thinking and serves as a starting point for the philosophical analysis of disease.

In spite of philosophers like Canguilhem (1991) pondering the issues concerning the divide between normal and pathological already in the 1940s, the concept of disease was still poorly defined in the sixties and in clinical discussions it was considered odd if one raised a question about what disease exactly is (Scadding 1963, p.1425). The term *disease* was in general use without a formal definition and most of those using it assumed everyone knew what it meant (Scadding 1967, p.877). Still in the 1970s, most medical textbooks did not attempt to define disease and most physicians seemed not too excited about the issue. The exact definition was perhaps considered unimportant or physicians might have thought they grasped it intuitively well enough. (Guze 1978, p.296.)

The current most influential naturalist accounts of disease (Boorse 1977 & 2014; Wakefield 1992 & 2014) are a result of a philosophical debate starting from controversies around psychiatry and the concept of "mental illness" in the 1960s and the 1970s. As the efforts of defining disease were in the beginning tightly intertwined with these controversies, it is probably good to start with an overview of them. The debate sparked interest in the definition of disease and contributed a lot to the change in understanding the meanings of "mental illness" and "disease" more generally. There were three particular events that

characterize the pressures for change and had a great impact for getting psychiatrists, physicians and philosophers interested in the subject of analyzing and defining the concept of disease.

The first one is not so much a single "event" as it is a movement which emerged in the early 1960s and lasted till the 1970s. Still, the birth of the *antipsychiatry movement* personified in Szasz (1960) may be viewed as the first of these significant events (Shorter 2005, p.22). The second event forcing psychiatrists to rethink their current conception of illness was the controversy concerning homosexuality in the early 1970s. The controversy revolved around the question of whether homosexuality should be included in the APA's[6] official nomenclature or not (Cooper 2015, p.85). The third event was the publication of Rosenhan's "Being sane in insane places" in 1973. Rosenhan's study was roughly about whether psychiatrists can distinguish sane people from the insane first in the auditions leading to potential admission and then in the hospital environment.

All these events had an impact on the discussions about the validity and definition of mental illness. As the concept of mental illness is deeply connected to the question about what illnesses or diseases fundamentally are, these discussions also had profound effects on the more general conception of disease. This is reflected in the works of Christopher Boorse, which are the primary focus of my research. Boorse was motivated by the mentioned developments in psychiatry but wanted to define the theoretical concept of disease so that it would apply primarily in the field of physical medicine but secondarily in psychiatry as well (Boorse 1975; 1976a & 1977).

Before proceeding further, it should be emphasized that the whole idea of "mental illness" as being fundamentally different from "physical illness" is highly controversial. The idea only developed towards the end of the 18th century and was based on the then popular Cartesian dualism (Kendell 2001, p.490) which has by now largely gone out of fashion, even though the mind–body -problem remains unresolved in psychopathology (Murray 2011, p.21). Kendell (2001, p.491) goes as far as claiming that the distinction (between disorders of mind and body) "has been long since abandoned by all thinking physicians".

Still, due to historical reasons and the current practical and nosological divide between psychiatric and somatic medicine, it is for my purposes reasonable to speak separately of

[6] American Psychiatric Association

the mental and the physical on certain occasions. And as we will see, the efforts on defining mental illness, and disease in general, have gone quite closely together.

2.2 The challenge of the antipsychiatrists

In spite of the long history of medicine, the need for clear boundaries between health and disease became imperative only in the 1960s when Thomas Szasz (1960) and other so-called "antipsychiatrists" published their critiques of the then current psychiatric establishment and its practices (Boorse 1976a, p.61; Kendell 1975, p.305; Wakefield 1992, p.374–375). The radical claim Szasz made was that there is no such thing as "mental illness". In Szasz's view, the conditions psychiatrists are dealing with are not "diseases" comparable to those of physical kind (Szasz 1960, p.116). These conditions are just ordinary social problems or problems in living labeled as "illnesses" and the demand for the labeling is driven by current norms and values (Szasz 1960, p.117). Szasz thought that the redefining of "problems in living" and disapproved behavior as illnesses was used as a way of justifying medical interventions to phenomena for which other means of intervening (i.e. social or political) would be appropriate.

Szasz's claims certainly were not groundless even if many disagreed with the implications. At that time, American psychiatry was dominated by Freudian psychoanalysts whose view was roughly that there is no clear line between health and illness and everybody has some degree of "mental pathology" (Decker 2013, p.3–4) in the form of "universal unconscious conflict" (Decker 2013, p.50). Under the influence of Freud and operating with the Meyerian biopsychosocial approach, they expanded psychiatry's focus beyond pathology and illness into social issues (Murray 2011, p.23). Many psychiatrists and even the most influential American psychiatric institutions – American Psychiatric Association (APA) and National Institute of Mental Health (NIMH) – explicitly took a stance according to which psychiatrists *should* deal with social problems (Decker 2013, p.4–5). Also, the DSM-II[7] (1968) included "nondisorder problems" in addition to the classification of "mental disorders" (Spitzer 2001, p.353).

Other antipsychiatrists typically agreed with Szasz that "mental illnesses" should not be considered as diseases, but their arguments for the claim varied. Mental illnesses were claimed to be learned behavior better dealt with by psychologists rather than psychiatrists

[7] Diagnostic and Statistical Manual of Mental Disorders II

(Eysenck 1960), rational ways of coping with pressures placed by societies and families (Laing 1960 & 1967), responses to the shocks of being labeled and treated as insane (Scheff 1963) or labels for justifying social control (Foucault 1965; Goffman 1961; 1963). Still, the common characteristic of all these critiques loosely associated with the "antipsychiatry movement" is the claim that the phenomena called "mental illness" – whether real or not – does not belong to the medical sphere and should not be primarily dealt with by physicians.

The antipsychiatry movement was quite popular in America and gathered a large following in the 1960s and the 1970s. Its success reflected the antiestablishment mood of the era, thriving due to the unpopular Vietnam War, racial tensions, sexism and other grievances the current administration was blamed for. The counterculture was not limited to hippies but could be found on college campuses among students and faculty. The antipsychiatrists' goal of undoing the supposed medicalization of social issues was embraced by both the left and the right ends of the political spectrum. Interestingly, Laing further widened the following by associating with people interested in the use of "mind-expanding drugs". (Decker 2013, p.23–24.)

In addition to the pressures placed by the antipsychiatrists, there was growing discontent also among psychiatrists themselves[8]. Biological psychiatrists – perceiving mental illness as a sign of physical brain disease – attacked psychodynamic thinking and derided psychoanalysis as a religion (Decker 2013, p.9). Meanwhile, a small group of scientifically-minded descriptive psychiatrists – later called the neo-Kraepelians[9] – were developing their own model of mental illness at Washington University in St. Louis (Decker 2013, p.53). These psychiatrists, such as Eli Robins, Samuel Guze and George Winokur, were dissatisfied with the current state of psychiatry. Like biological psychiatrists, they thought the dominant Freudian model based on psychoanalytic concepts was unscientific and its etiologies[10] pure speculation. They also argued that psychiatry was dealing with nonpsychiatric pursuits (i.e. social phenomena) while it should stick to the medical model and make a clear distinction between health and disease.

[8] Some antipsychiatrists were also professional psychiatrists themselves. Szasz and Laing for example.
[9] Named after Emil Kraepelin (1856–1926), "the father of modern psychiatry"
[10] Etiology = theory or study of causation of disease.

When people are saying that psychiatry "should stick to the medical model", they typically mean that psychiatry should apply a similar model that is applied in physical medicine. "Medical model" is a not a well-defined concept though and it is given different meanings by different writers (Boorse 1976a, p.62; Guze 1978, p.295). It usually involves certain beliefs about the nature and causes of disease (Boorse 1976a, p.62; Guze 1978, p.295; Kendler & Zachar 2007, p.560), usage of medical vocabulary, support of medical interventions done by physicians and preferring the use of drugs over therapy (Boorse 1976a, p.62). Roughly, the point of emphasizing the medical model in the context of psychiatry is to give support to the idea that "mental disorders" are just a subset of medical disorders (Spitzer 1978, p.657) and should be addressed by similar medical means.

As the knowledge about the nature of psychiatric illness was rather limited, the neo-Kraepelians thought that for the time being psychiatry should be centered upon *observation* and *description*. This kind of descriptive psychiatry would make the communication between professionals better and boost progress in research. (Decker 2013, p.53–54.) The neo-Kraepelians later came to play a significant role in the change of paradigm in psychiatry and in the development process of DSM-III (1980), the revolutionary classification of psychiatric illness.

As we can see, the view that psychiatry had unjustifiably widened its scope to social matters was shared between the neo-Kraepelians of St. Louis school and the antipsychiatry movement. But the two groups were separated by their conclusions: the former thought that psychiatrists needed to sharpen their game and stick to the medical model whereas the latter advocated abandoning the medical model and addressing "mental illness" as a form of social phenomena by sociopolitical means. Yet both were motivated to get rid of the current establishment even if their ultimate goals were not aligned.

For now, I do not intend to delve deep into the discussions between the antipsychiatrists, the neo-Kraepelians and their opponents. I just want to note that the motivation in defining "illness" or "disease" rose in this context of partly sociopolitical and partly academic controversy[11]. The arguments of the antipsychiatrists were problematic for the psychiatric establishment if it wanted to maintain the status of scientific medical discipline and the prestige that comes with it (Decker 2013, p.xvi). To maintain the status and professional

[11] There was additional pressure from public and private third-party payers (e.g. insurance companies) which wanted a clear categorical distinction that could exclude people from therapy rebates (Murray 2011, p.24).

credibility, the proponents of the medical model in psychiatry[12] had to show that mental illness is a phenomenon really belonging to the medical sphere. To do that, a valid definition was needed; a definition that would set the boundaries between health and illness and thus limit the phenomena psychiatrists should be dealing with. Since the dominant Freudian psychoanalysts in the 1960s and 1970s considered health and illness as ends of the same continuum where all mental life ranged (Decker 2013, p.3), they were ill-prepared for the task. This – in part – paved the way for the neo-Kraepelians and also philosophers like Boorse, who was a proponent of making the boundaries clearer.

2.3 Is homosexuality a mental disorder?

Along with the antipsychiatry movement, the final push for reforming the concept of disease came from two directions. First, there was the controversy around the medical status of *homosexuality* in the early 1970s (Spitzer 1978, p.657). During that time, homosexuality was still part of the American Psychiatric Association's official nomenclature DSM-II, i.e. homosexuality was considered a *disorder*. The diagnosis was removed from the manual only in 1973 (Boorse 2011, p.13; Hinderliter 2015, p.128). The removal was strongly influenced by gay activists' campaigning and their representatives' meetings with APA's Nomenclature Committee. The activists' claim was that psychiatric profession was one of the "gate-keepers" of society's attitudes and removing homosexuality from the list of mental disorders would have profound effects on the lives and rights of gay people. (Silverstein 2008, p.161.)

Gay activists' protests were dated at the height of the antipsychiatry movement and sometimes the situation got quite dramatic. In 1970 gay-activists joined by other protestors with antipsychiatry stance attacked APA's annual convention and formed a human chain around the convention center preventing psychiatrists from entering the meeting. There was shouting and foul language from both sides and one physician even called for the police to shoot at the protestors. (Cooper 2015, p.85; Decker 2013, p.31.) This depicts quite well how inflammatory the situation was around psychiatry in America in the 1970s. There was massive socio-political pressure for change in both psychiatric practices and nomenclatures.

[12] See Guze (1978) for a defence of medical model in psychiatry at the time of the controversy.

Finally, the APA gave in and the eventual removal of the homosexuality diagnosis from the DSM in 1973 meant that APA had to thoroughly rethink certain aspects of their disorder-concept and classification to maintain consistency and credibility. It was thought that the removal could be rationalized by adding a "harm-criterion" to the diagnostic criteria. Consequently, a sexual deviance like homosexuality would be an illness only if it caused significant harm or "distress" to the individual concerned (Cooper 2015, p.85; Spitzer 1978, p.660). Thus, homosexuals content with their sexual orientation would not be diagnosed.

The problem was that there were many other diagnoses in the DSM which would also be compromised under the new harm-criterion, e.g. *high-functioning autism* or *low I.Q.* (Cooper 2015, p.83–84). It is plausible that there are reasonably happy autists and low I.Q. people with no subjective distress. Also, there were diagnoses for "sexual perversions" like necrofilia for which there might not be subjective distress involved[13]. If these kinds of diagnoses were to be kept in the classification, a definitional harm-criterion would be problematic. At the end, the proposed strict harm-criterion was down-graded to "typical association with distress" in the published version of the DSM-III (1980). (Cooper 2015, p.86.) The final definition of mental disorder in the DSM-III turned out hardly satisfying and the challenge of constructing a definition which would not be over-inclusive or over-exclusive remained.

2.4 Insane people or insane places?

The second individual event piling up the pressure for change was the publication of the article "On Being Sane in Insane Places" from David L. Rosenhan in 1973 by the prestigious journal *Science*. The article was about Rosenhan's experiment in which he got eight "sane" people (including himself) secretly admitted to twelve different mental hospitals as "insane", i.e. with a psychiatric diagnosis (Rosenhan 1973, p.251). These "pseudopatients" had no history of mental illness and they represented a diverse group of men and women of different occupations. The presence of the pseudopatients and the nature of the research program was not known to the hospital staffs. (Rosenhan 1973, p.251.)

[13] Or if there was distress, it might be experienced by someone other than the "pervert" himself.

After the admission to the ward, all pseudopatients were instructed to act "normally" and the idea was to find out whether hospital staff would detect them as "sane".

Rosenhan's conclusion from the experiment was roughly that the failure of hospital staffs to detect the pseudopatients showed that psychiatric professionals cannot distinguish the sane from the insane. The original diagnosis in the admission was made based only on false symptom-descriptions and then in the hospital ward the normal behavior of pseudo-patients was misinterpreted as insane due to environmental effects and labeling[14]. (Rosenhan 1973, p.257.)

If Rosenhan's conclusion was correct, it would present a great embarrassment to the psychiatric establishment by making its diagnostic and treatment practices seriously compromised. Surely, Rosenhan's methods and conclusions were vigorously criticized (see Guze 1978; Millon 1975; Spitzer 1975). But even if there were problems with Rosenhan's methods and experimental arrangements, he still managed to highlight certain serious shortcomings of psychiatric practice (Millon 1975, p.456) in a way that appealed not only to the antipsychiatrists and their followers but also to many psychiatrists. The great publicity the article got contributed to the pressure towards the psychiatric establishment to address not only the practical matters associated with diagnostic and clinical work but also the theoretical issues concerning their classifications and conception of illness.

Because Rosenhan's experiment was interpreted by many as confirming the social nature of "mental illness", it motivated the naturalists in their pursuit of showing that *disease* is fundamentally a natural, scientific and value-free concept. To succeed, they needed to construct a new definition of disease which would exclude social values.

The work of the most influential naturalists on disease today – Boorse (1977 & 2014) and Wakefield (1992 & 2014) – is clearly motivated by these developments and events in psychiatry that I have presented in this chapter. However, there were naturalists before them, the most significant ones (in this context) being Scadding (1959; 1963 & 1967) and Kendell (1975 & 2001). Of these two, Kendell was also motivated by the controversies revolving around psychiatry but Scadding was probably not. Scadding published his first relevant papers before the homosexuality controversy and Rosenhan's experiment and

[14] "Labeling" refers here to the idea that psychiatric diagnosis sticks with its bearer, stigmatizes him and guides the interpretations made of his behavior. Labeling theory is presented by Scheff (1963) and Goffman (1963) just to name a few. Guze (1978, p.301) notes that people not believing in "mental illness" often claim that psychiatric diagnosis is merely a way that society "labels" its deviants.

did not make any reference to psychiatry in his texts. Scadding's work (with the help of Kendell) laid the foundations on which Boorse started to build his Biostatistical Model. Hence, I will start chapter 4 about the Biostatistical Model by presenting first Scadding's and Kendell's ideas on disease and then proceed to analyze the BST.

But before that, I want to make a few remarks about conceptual analysis and elucidate the philosophical positions relevant to my analysis.

3 Conceptual analysis, naturalism and revisionism

Before proceeding into the details of Boorse's Biostatistical Model and the analysis of health and disease, it is worthwhile to take a glance at the *method* Boorse uses in his work and which I also make use of throughout this essay. This method is *conceptual analysis* widely applied in philosophical inquiry.

In fact, some people think that *philosophy is essentially the a priori analysis of concepts*. This position is controversial though and the status of conceptual analysis divides philosophers deeply (Margolis & Laurence 2014). Still, the method is common in philosophical texts and my purpose is not to challenge it here. Rather, my aim is to assess how conceptual analysis is approached and what its goal is in the context of Boorse's work.

In this chapter, I will also elaborate what it means to be either a *conservativist* or a *revisionist* about concepts. Lastly, I explain *naturalism* and *normativism* in the context of philosophy of medicine and psychiatry.

3.1 The goal of conceptual analysis

What kind of an answer one is after when asking "what is disease"? Kingma argues there are at least three possibilities:

1. One wants an exact description of all the ways in which people use the term *disease* and all the things people are (or think they are) referring to with that term.

2. One wants to discover the real-world category that the term *disease* picks out. For example, in chemistry it was discovered that the term *water* picks out the natural category H2O.
3. One wants a mix of the first two options. One studies how people employ the term *disease*. But one is not satisfied in only reporting his findings. One assumes that people are often messy and inconsistent in their use of language and concepts. One's task lies in uncovering the underlying general ideas and features that some way unite our messy common usage of the concept of disease. This project relies on actual usage of the concept, but unlike the first option, intends to revise and refine it. This third option might in some cases generate the same result as the second option, but the difference is that the third option does not *presuppose* that the concept picks out a natural category. (Kingma 2017, p.48–49.)

The third option describes roughly what is usually called *conceptual analysis*. That is also what philosophical work with concepts mostly seems to be about. Boorse's project of analyzing the concept of disease fits into this description quite well. Boorse is a naturalist in a sense that he thinks diseases are real-world entities which exist independently of our values or attitudes towards them, but his analysis does not *presuppose* this. Boorse studies the usage of *disease* (or pathological condition) in medical texts, makes an analysis of the data and according to that analysis refines and revises the concept so that the final product reflects what he sees as the concepts "correct meaning". After all this conceptual work, he concludes that yes, *disease* in fact picks out a real-world category of certain kind of biological conditions. But he arrived at this conclusion via analysis, the analysis did not presuppose it – even if Boorse himself sometimes seems to presuppose it. This methodological point is good to keep in mind especially when trying to assess the validity of different criticisms.

Haslanger & Saul elaborate further what forms of conceptual analysis there exist. They distinguish between three different approaches on analyzing concepts: *internalist, descriptive* and *ameliorative*. Following an *internalist approach* in the case of "disease", one would be asking "what is our concept of ´disease´?" and then relying on *a priori* methods for an answer. On a *descriptive approach*, one asks what objective types our

epistemic vocabulary tracks. For example, naturalists may start by identifying paradigm cases of disease and then work on to identify the relevant kind to which the paradigms belong, drawing on empirical (or quasi-empirical) research. Finally, *ameliorative approaches* are about the purpose and meaning of concepts. Following an ameliorative approach, one may ask why we have a concept of "disease" and what kind of concept would work best for our purposes. (Haslanger & Saul 2006, p.94–96.)

As Haslanger & Saul (2006, p.96) note, these different approaches are not entirely distinct. This becomes clear in my following analysis of the debate about health concepts. For example, Boorse explicitly tries to stay on the descriptive side while constructing the Biostatistical Model of disease. He intends to track medical usage of *disease* and explicate the true meaning of the concept as applied in medical theories. Nonetheless, Boorse quite frequently deals with questions concerning the fundamental nature of disease and the point of having the concept in the first place. Hence, it would be misleading to conceive Boorse's project purely as descriptive conceptual analysis.

It was already hinted that one thing Boorse does with his analysis is *revising* the concept of disease. This gets us smoothly to the second point that deserves a brief discussion before proceeding further. It is about the question of a degree in which there is value in preserving the common-sense meaning of concepts when creating scientific/academic language.

3.2 Conservativism vs. revisionism

When applying conceptual analysis as a method, there is one important decision to be made. The decision concerns the degree in which one cares about whether the target concept under construction agrees with the intuitions behind the concept's use in common language.

If one thinks it is crucial that scientific concepts and common-sense concepts agree, one is called a *conservativist*. In this case, one thinks there is value in preserving most of the meaning of the concept commonly in use. If one does not care much about whether these different types of concepts agree at all, one is called a *revisionist*. A revisionist may think that in science, all that matters is that the concepts are defined in a way that best serves scientific practice. Common-sense constraints might hinder scientific development. But

even a revisionist usually must use a common-sense concept at least as a starting point of his analysis. (Murphy 2015.)

To be able to analyze a concept, it must either already exist in some linguistic context or be made up in the analyzer's head. Usually, the concepts to be analyzed exist in common language or in academic literature. Sometimes – perhaps in the cases of creating completely new philosophical or theoretical physics' concepts – there is no existing concept where to start from. Then the distinction between conservatism and revisionism might be inconsequential. But the usual case is that the analysis starts from a pre-existing concept.

Conservativists and revisionists may also be distinguished by their tendencies towards naturalism. There are conservativist and revisionist naturalists as well as there are conservativist and revisionist constructivists[15]. A *revisionist naturalist* argues that we must go with our conceptual analysis where the science leads us even if that means we eventually end up using language in ways that look bizarre from the common-sense perspective. In contrast, a *revisionist constructivist* may argue that the reasons for revising and refining our concepts are connected to some social goals, such as emancipation of oppressed groups. For *conservativist naturalists*, folk concepts specify what health and disease are. Then the job of medical science is to find out whether anything in nature falls under the concepts revealed by analysis. Interestingly, Murphy notes that naturalists tend towards conceptual conservatism[16]. Constructivists – on the other hand – are usually revisionists who want to change our language in a quest for social change. A *conservativist constructivist* would argue along the lines that "our folk concept of disease is that of a pattern of behavior or bodily activity that violates social norms", without an intention to mold the language. (Murphy 2015.)

Above I have explained how conservatism and revisionism can be conceptualized as *attitudes* or *stances* towards the relationship between scientific and common-sense concepts. These attitudes and stances are derived from beliefs about the potential value of retaining the meanings of concepts currently in use. But the distinction may be assessed from a more practical perspective as well. Someone might do conceptual analysis without any kind of attitude or stance concerning conservatism and revisionism. Still, an *observer* might call the analysis in question conservativist or revisionist after getting acquainted

[15] For some reason, Murphy (2015) uses "constructivist" to refer to similar position that is usually called "normativist" in the literature.
[16] Which clearly is not the case with Boorse.

with the work done and the resulting concept (e.g. Kingma 2013, p.377–378 analyzing Wakefield's (1992) model). If the concept resulting from the analysis is close to its common-sense counterpart, then the analysis can be called conservative, whether the analyzer has any personal preference for conservativism or not. Similarly, if the resulting concept differs significantly from the common-sense one, the analysis can be called revisionist.

I suggest that the above may be framed as a distinction between *subjective-prescriptive* and *objective-descriptive* perspectives. From the subjective-prescriptive perspective, the question of conservativism vs. revisionism is about beliefs and preference: one's preference for the former or the latter depends on how he believes the costs and the benefits of either option will balance out. From the objective-descriptive perspective, the question is just about how things are: in what degree the observer interprets a given theoretical concept and a given lay concept agreeing in their meanings.

Boorse is a strict *revisionist naturalist* at least in the first sense. He does not care whether the medical concept of disease he defends agrees with the common-sense concept of disease. But whether (and to what degree) his analysis manifested in the BST is revisionist in the second sense is not obvious[17]. Even though Boorse does not care about following common intuitions about disease, his analysis is meant to describe theoretical medicine's use of the concept as it is (purified of misunderstandings and messy use of language). The question is how much the theoretical concept resulting from his analysis happens to deviate from the common-sense concept.

One problem making the assessment difficult is that – as noted earlier – there are many terms referring to "problematic medical conditions" in both common language and medical language. In fact, and unfortunately, this messy use of terminology sometimes plagues even the philosophical literature I refer to throughout this thesis. On the bright side, Boorse is always very precise. But in both common language and medical language generally, the terms used to refer to "problematic medical conditions" include at least *disease*, *illness* and *disorder*. Thus, even if Boorse manages to correctly analyze the medical and theoretical "disease", it is not easy to say how close the result of the analysis would be with the common-sense concept. To make this assessment, one would first need

[17] Murphy (2015) argues that because Boorse's model relies on "functional decomposition as the ultimate justification of judgments of health and disease", it must be a revisionist account in this second sense.

a clear understanding of the meanings of *disease*, *illness* and *disorder* in common language. Do any or perhaps all of them refer to the similar concept in common language as *disease* does in medical theory? How similar? These are open questions.

Fortunately, as Boorse is a revisionist in the first sense as I presented, the potential difficulties in assessing how his disease-concept *de facto* corresponds to its common-sense counterpart do not demonstrate a major weakness in his model[18]. If Boorse cared about the equivalence of the theoretical and common-sense concepts and claimed that his concept of disease corresponds to the common-sense one, the factual non-correspondence would clearly hurt the model and his claims. But now that he is an explicit revisionist thinking the benefits of his approach outweigh the costs, a potential observation that his disease-concept does not match with common-sense is of minor significance.

Furthermore, now that Boorse has replaced *disease* with *pathological condition* (Boorse 2014, p.684) in his model, it is not obvious whether this new term or the concept it refers to even have established positions in common language. Many people probably have heard of *pathological condition*, but my guess is that only a small minority has more than a vague feeling of the concept the term refers to. If it is the case that *pathological condition* is not an established term in common language, it is probably less bad to deviate from its meaning in medical language – even from a conservativist point of view. The costs of deviating from an established term's meaning should be larger than the costs of deviating from the vague meaning of an ambiguous term.

Still, we are left with the question about whether the BST is "revisionist" in relation to the existing *medical concept of disease*, i.e. whether the BST tracks the medical usage of *disease* closely enough or not. Boorse would say the BST tracks the "correct" medical usage as closely as possible (1977, p.551), since that is what the model is designed for. Some would disagree (see Boorse 1997, p.41; 2014, p.710), but the question is up for debate.

The point of my analysis in this section is that when trying to assess Boorse's model's potential merits or drawbacks, it is essential to keep in mind what he is trying to do.

[18] This means only that as an explicit (attitudinal) revisionist, Boorse may legitimately ignore the problem of assessing potential disparities between his theoretical concept and the common-sense concept. But it does not mean that the disparities could not be a problem if it turns out that revisionism is problematic in this context, i.e. if there happens to be large social or other costs due to theoretical and common-sense concepts not agreeing.

Boorse is primarily doing *conceptual analysis* which aims to track the correct usage of *disease* as displayed in theoretical medical texts. The "correct" signals that he means not just to report all the different and perhaps sometimes confused ways in which the terms are used. Boorse means to define the concept so that it best resembles – by his interpretation – the one which medical theorists have in mind when they write about disease. He is not trying to construct the concept from scratch or trying to align it with common-sense.

Hence, when criticizing the BST, the critic should distinguish whether he is 1) disputing the model as the correct analysis of medical usage of disease or 2) agreeing about the analysis but claiming that medicine's concept revealed by the correct analysis is flawed one way or the other. Furthermore, the BST should not be criticized for not agreeing with common sense *per se*. If one wishes to criticize theoretical concepts not agreeing with lay concepts, one should provide a general criticism of the revisionist position. Or if one thinks that in this particular case it is especially problematic that the medical and lay concepts do not agree, one should provide an argument for that claim. But only noting that the BST does not agree with common-sense does not tell anything beyond the obvious.

3.3 Naturalism and normativism

Now I will explain what is meant by naturalism and normativism in the context of philosophy of medicine and psychiatry and especially in the context of discussion about health and disease. This short description aims not to give an exhaustive analysis of naturalism and normativism as *metaethical positions*. The aim is just to give an idea of the meaning and use of these concepts in the context relevant for my essay.

Naturalists view "disease" as an empirical, value-free concept that picks out a real-world category (Kingma 2017, p.49) that exists independent from our values and interests (Kingma 2013, p.364). If health is understood in a traditional medical way as the *absence of disease* (Boorse 1977, p.544), then health is an empirical and value-free concept too – from a naturalist perspective. Typically, naturalists assume that what the category "disease" picks out in the nature is "biological dysfunction" (Kingma 2017, p.49). The con-

cept of biological dysfunction presupposes that there is a way of functioning that is standard or "normal". "Normal functioning" is then equated with "health" (Boorse 1977, p.555). Thus, "disease" is a deviation from normal functioning, i.e. "dysfunction"[19].

Naturalists maintain that norms or values do not determine what dysfunction is. Dysfunction is determined by natural science. On the naturalist view then, the history of psychiatry – for example – reflects both mistaken science and illegitimate social norms affecting disease judgements. (Kingma 2013, p.365.) If norms or values are affecting what is considered a disease, it is always at least somewhat problematic from a naturalist perspective, whatever those norms or values happen to be. According to naturalists, it is *science only*, which should determine what dysfunction and thus disease, is[20].

The most influential naturalist accounts of disease today are Christopher Boorse's (1977; 2014) Biostatistical Model (BST) and Jerome Wakefield´s (1992; 2014) Harmful Dysfunction Analysis (HDA) (Kingma 2013). Of these two, Boorse's account is archetypally naturalist and the one normativists and other critics have attacked the most (Kingma 2017, p.53). Wakefield's evolutionary account is naturalist in a sense, because it has *dysfunction* as a necessary condition for disease (Kingma 2013, p.372; Wakefield 2014, p.649). The fact that *harm* is also a necessary condition in Wakefield's account and that the two necessary conditions are sufficient for disease only together, make some people call the account *hybrid* (Wakefield 2014, p.650). Ironically, Boorse (1997, p.5) calls Wakefield a *normativist*, because of the harm-condition[21]. Thus, Wakefield's account may be viewed as naturalist, hybrid or normativist, depending on which aspect the observer wishes to emphasize. But crucially, as *dysfunction* is necessary for disease in Wakefield's account, much of the analysis targeted at the BST also apply to the HDA.

Normativism is a view that health and disease are essentially evaluative concepts (Kingma 2017, p.54). According to normativism, there is no natural, objectively definable set of biological malfunctions that cause disease. Rather, normativists assert that to call a condition a disease is to make a judgement that someone in that condition is suffering a specific kind of harm that we explain in terms of biological processes. (Murphy 2015.)

---

[19] More on functions in chapter 4 about the Biostatistical Model.

[20] It is debatable whether natural science itself is value-free or not. Naturalists in the context of health concepts usually take it as given that natural science is value-free *enough* (Boorse 1997, p.56).

[21] Broadbent asserts it is misleading to describe Wakefield's HDA as a *hybrid*, because there is no logical space between health facts being value-laden and value-free. He sides with Boorse and prefers *weak normativist* instead of *hybrid* for accounts like the HDA. (Broadbent 2017, p.3.)

While naturalism is quite straightforward to describe, the sense in which normativism is applied in different contexts varies and maybe difficult to grasp. Normativists share the view that health and disease are value-laden concepts, but do not necessarily have a shared understanding of value-ladenness and its meaning. They might also have different conceptions about the metaethical status of values. (Kingma 2017, p.55.) Thus, just noting that someone is a normativist does not say much about the *way* one sees health and disease are value-laden or about what is one's metaethical position on values.

It is also important to be careful not to confuse people *describing* the current concepts of health and disease as value-laden with people holding either a normativist position according to which health and disease *cannot be* given a purely naturalistic definition or a normativist position saying that health and disease *should be* normatively defined. A naturalist may judge an existing concept as evaluative but still hold that it can and should be defined value-free.

For example, Szasz (1960) was a *naturalist* claiming that psychiatry's concept of "mental disorder" is evaluative. Szasz thought that diseases can and should be defined naturalistically as physical lesions. As there are no signs of physical lesions in almost any mental disorders, Szasz claimed they cannot be *real diseases*. Szasz provided an argument explaining the way he thinks mental disorder is an evaluative concept and why he thinks it is bad to use such an evaluative concept of disorder within a discipline that claims to be scientific and a branch of medicine. (Kingma 2013, p.366.)

Normativists could agree with Szasz that "mental disorder" is a problematic concept, but the reasoning behind the assessment would be different. They *could not* say (like Szasz) that the concept of mental disorder is flawed because it is not based on natural facts. They *could* say that it is problematic, because the values behind the concept are "wrong" and should be replaced with better values, then perhaps leading to different set of conditions viewed as disorders. Or they could say that the values are fine, but because mental disorder is value-laden, it should not be promoted by psychiatry as a purely scientific medical concept. Perhaps mental disorders should be dealt with by social workers, changes in social policies or some other measures other than medical interventions. Finally, if a normativist thought that the values behind "mental disorder" are just fine and the concept is well applied in psychiatry, he could conclude that Szasz is wrong in two ways: first, he is wrong to think that diseases could or should be defined in a value-free way, and second,

he is wrong to think that the evaluative character of "mental disorder" would imply that psychiatrists should not deal with mental disorders.

More generally, what normativists might claim is that health and disease cannot be defined in such a way that the definitions themselves are not overtly or covertly value-laden or that if they *are* defined in such a way, then most of the meaning is lost in the process (e.g. Fulford 2001). Even more typically, they might claim that viewing diseases as "bad conditions" is such an essential element of our understanding of disease, that it cannot or should not be taken out of the concept's definition (e.g. Cooper 2002 & Reznek 1987).

Nordenfelt has constructed a more revisionist normativist account, which defines "disease" as a second-order inability to reach one's vital goals. On this view, one's vital goals are the goals that are jointly necessary and sufficient for minimal happiness. A second-order inability is the inability to gain an ability. (Kingma 2017, p.58.) What's normativist about this account, is that it defines "disease" in relation to reaching one's "life-goals" and happiness, which both are evaluative concepts.

The standard problem for normativist accounts of disease is that they have not been able to explain why we routinely make distinctions between pathological conditions and those conditions we just disapprove of. We might disapprove or consider "bad" many things, such as selfishness, impoliteness or ugliness. Still, we mostly do not consider selfish, impolite or ugly people as diseased. Diseases seem to form a special subclass of "bad conditions". The normativists need to explain what kind of a special class of value judgements are those judgements, which single out diseases among all bad conditions and why diseases are candidates for a particular set of causal explanations. In contrast, a naturalist may simply answer that diseases are those (bad) conditions which involve biological malfunctions. (Murphy 2015.)

Perhaps in addition to some Neo-Aristotelian accounts, the accounts of Cooper, Reznek and Nordenfelt are considered the most important normativist accounts in the literature (Kingma 2017, p.56). I mentioned their main ideas here just to give a taste of what kind of claims normativists promote. I won't go into the details of normativist accounts as they are not that relevant for my purposes. For the details, see the mentioned works of Cooper (2002), Reznek (1987) and Nordenfelt (1993; 2017). For criticism, see Kingma (2017, p.55–56).

4 The Biostatistical Model

4.1 The foundation of the Biostatistical Model

In physical medicine – even though there wasn't any generally accepted definition – throughout the latter half of the 19th century till around 1950s there had been a widely adopted conception of diseases as manifestations of *physical lesions*[22]. This conception was due to the developments in morbid anatomy and histology, both of which provided evidence of structural damage associated with diseases either at gross or microscopic level. At the time this evidence seemed to suggest that an *identifiable lesion* should be regarded as the essential attribute of *disease*. (Kendell 1975, p.308.) This conception started to wobble in the 1950s and finally crumbled in the 1960s after Pickering et al. (1960) had shown how a common disease like hypertension was in fact graded in character and shading insensibly into normality (Kendell 1975, p.309). It was then understood that not all disease could be distinguished from health qualitatively by showing the existence of a physical lesion.

After discarding the concept of disease based on a physical lesion it seemed likely that its successor would have to be a statistical model which would define the normal range of variation in human traits (Kendell 1975, p.309). But as people are diverse and differ greatly in many kinds of traits and most deviations are obviously not problematic or harmful, a statistical model which *only* defines the normal ranges of variation is not enough. Criteria for distinguishing between *harmful deviations* and *those deviations which are consequentially neutral* (or positive) was needed. J. G. Scadding was the first to formulate a definition along these lines (Kendell 1975, p.309) and I will introduce it next.

Scadding coined his definition of disease first in a 1959 paper in The Lancet and refined it later (1963 & 1967). I will present here his refined definition. It conveniently highlights both the essential parts of an advanced kind of a new definition that was (supposedly) needed to replace the old lesion-based account:

> A disease is the sum of the abnormal phenomena displayed by a group of living organisms in association with a specified common characteristic or set of characteristics by which they differ from the norm for their species

[22] *Lesion* refers to a recognizable deviation in organism's anatomical structure (Wakefield 1992).

> in such a way as to place them at a biological disadvantage (Scadding 1967, p.877).

By the "set of characteristics by which they differ from the norm" Scadding implies a statistical concept of normality and abnormality. The criterion that the deviation from the norm should "place them at a biological disadvantage" is meant to distinguish deviations which should be regarded as associated with disease from those which should be not. This biological part of the definition is what makes Scadding a *naturalist*. It should be emphasized that Scadding makes no effort to explain what exactly his concept of "biological disadvantage" refers to. He just acknowledges it is a vague term and goes on with his analysis. This disregard may be due to his view that a "unified (general) concept of disease" is impossible and instead there are many different kinds on defining characteristics for different diseases (Scadding 1959, p.323; 1963, p.1425–1426; 1967, p.877). For Scadding, the purpose of his *general definition of disease* was just to make the communication between professionals clearer and the important definitional work should be done in the context of individual diseases.

Kendell then took on the task of elaborating the meaning of "biological disadvantage" in Scadding's definition. According to Kendell (1975, p.310), "it must embrace both increased mortality and decreased fertility". Rephrased, biological disadvantage refers here to *decreased survival or reproduction* (compared to the norm). Kendell does not really explain or justify the choosing of survival and reproduction as the goals of biological organisms and this lack has generated lots of criticism (see Boorse 1997, p.23 & 2014, p.691–694). Whether these stated goals were problematic or not, making them explicit contributed to the further development of Scadding's original ideas.

Kendell thought Scadding was on the right track with his statistical approach and the biological disadvantage -criterion was needed to limit the inclusiveness of the definition (Kendell 1975, p.310). Not every harm should be treated by physicians – only those that have biological origins and may reasonably be relieved by medical means. The biological criterion of this kind would help drawing the line between "real diseases" and "social problems" which, e.g., antipsychiatrists were concerned about. This reflects the basic naturalist position which assumes that ultimately disease is a scientific biological phenomenon.

A further benefit the biological disadvantage -criterion provides for the definition of disease (in contrast to the old lesion-based view) is that it accepts the significance of environmental influences and their complex interactions with individual properties in disease processes. Conditions like sickle cell trait or albinism – which could be harmful or beneficial depending on the current environment – would be considered diseases only in environments in which their presence is a real disadvantage. (Kendell 1975, p.310.) This seems to be a nice property to have for the concept if it is aimed to track those conditions people care about[23]. The lesion-based definition could not take environmental variation into account as its only basis for demarcating disease was the presence of a lesion.

Scadding's and Kendell's work laid the foundations for the new naturalist and statistical disease concept. It was the basis which psychiatrists and philosophers started to build on when faced with the pressures for delineating their area of expertise and defining disease.

4.2 The model defined

Now, I will present Christopher Boorse's Biostatistical model (BST) and analyze its key elements. As we will see, Boorse owes a lot to his predecessors, mainly Scadding and Kendell, who were the pioneers in combining naturalistic and statistical approaches together. I will start from the model itself and then delve deeper into the details. The BST has not changed much since its first formal introduction in 1977, but here it is probably appropriate to use the most recent version from 2014. So, this is how Boorse currently defines health and disease (or pathological condition):

> 1. The *reference class* is a natural class of organisms of uniform functional design; specifically, an age group of a sex of a species.
>
> 2. A *normal function* of a part or process within members of the reference class is a statistically typical contribution by it to their individual survival [or] reproduction.

[23] It is not clear if this is as "nice" for a Boorsian definition of disease which is meant as strictly theoretical and downplays practical qualities. Boorse targets medical – or "pathologist's" – usage with his concept, not so much things people care about (Boorse 1997, p.25 & 2014, p.683–684).

> 3. *Health* in a member of the reference class is *normal functional ability*: the readiness of each internal part to perform all its normal functions on typical occasions with at least typical efficiency.
>
> 4. A *disease* [later, *pathological condition*] is a type of internal state which impairs health, *i.e.*, reduces one or more functional abilities below typical efficiency. (Boorse 2014, p.684.)
>
> The following clause was added in 1997 to meet the challenge of atypical environmental diseases:
>
> [A] statistically species-subnormal function (in the usual sense of an arbitrarily chosen lower tail) is pathological if it results from an environmental factor outside an arbitrarily chosen central statistical range of that factor in the environments where the species lives. (Boorse 2014, p.684.)

Boorse's definition is then, as we can see, *functional* and *statistical*. There does not need to be any kind of physical lesion present for a condition to fulfill the criteria of *disease*: it is enough to show that there is a reduced function in some relevant part of the organism such that its contribution to survival or reproduction of the organism falls outside (species, sex and age typical) normal variation[24]. Boorse thus follows Scadding by adopting a statistical model of disease, handles the vagueness of Scadding's "biological disadvantage" by making it explicit what kind of disadvantage it is in question and replaces the old assumption about the necessity of a physical lesion with his functional account.

The words in brackets mark the changes from the 1977 version. In the older version, there was "and" instead of "or" in the definition of normal function. That led to confusion about whether Boorse meant that to consider a function normal, it should be a typical contribution to *both* survival and reproduction or that being a typical contribution to one or the other is enough. Boorse later clarified that he meant the latter and now the "or" signals that typical contribution to either goal is sufficient for normal function (Boorse 2014, p.685).

[24] Whatever the boundaries of normal variation are. Boorse remains deliberately ambiguous about this question. He thinks the boundaries of normal variation should be decided case by case depending on the disease and the reference class. He asserts that the boundaries can only be conventionally chosen, as in any application of statistical normality to a continuous distribution. (Boorse 1977, p.559 & 2014, p.714.)

The second change is the replacement of *disease* with *pathological condition*. This is due to Boorse's assessment that pathological condition is better in line with medical writing (Boorse 2011, p.26). The change is insignificant as it does not alter the details of Boorse's analysis.

The clause about "atypical environments" is needed to explain why conditions such as heat exhaustion are diseases by the BST, even if they are statistically normal in certain environments. The BST states that a normal function is a statistically typical contribution by a part or a process to individual survival or reproduction. But in atypically high-temperature environments heat exhaustion is a statistically normal reaction of humans. So, the BST without a clause about atypical environments would judge heat exhaustion as a *normal condition*. This problem cannot be dodged by deciding to use pure statistical species-typicality, ignoring completely the environment and what the individual is doing. Pure statistical species-typicality would resolve cases like heat exhaustion but would lead to conditions like high heart-rate during exercise to be considered as diseases. Thus, Boorse has decided to introduce atypical environments -clause which says that there are certain rare environments where humans typically do not live in, which may cause pathological conditions which are statistically typical in that rare environment. (Boorse 1997, p.83–84.) This makes the BST more complicated, but in principle the clause may achieve what intended. In practice, it may be difficult to draw the line between "typical" and "atypical" environments.

It is remarkable how the Biostatistical Model has remained mostly unchanged yet still relevant for almost 40 years despite all the criticism it has generated, especially considering that Boorse's original model in the 1970s was basically the first serious attempt to construct a general definition of disease in a somewhat rigorous way. There still aren't any notable naturalistic competitors around (Kingma 2017, p.52). Perhaps some would count Wakefield's evolutionary model (Wakefield 1992) as one, but I think – at least in the context of conversations concerning the value-ladenness of health concepts – Wakefield's model is more a complement than a competitor since it shares the relevant functional ideas with the BST. In its naturalistic details Wakefield's account differs notably from Boorse's only in what it counts as a function (Kingma 2013, p.372; Wakefield 1992, p.378). As disease judgements are based on departures from normal function, the analysis is largely similar. The fact that Wakefield adds a normative "harm criterion" to his model

(Wakefield 1992, p.381) might be interesting in many ways, but has no effect on the analysis of the separate naturalistic part of the account.

4.3 Key terms

Next, I will give a brief analysis of the key terms in the BST. The key terms are *function*, *reference class*, *health* and *disease*.

4.3.1 Function

There is a large philosophical literature on the problem of analyzing biological function statements (see Garson 2016). Boorse has argued for his own account – which he still applies in the BST – in detail (1976b), but here I will just outline the relevant features for my purposes, which are connected to the questions about value-ladenness.

In Boorse's view, functions are contributions to goals, which are derived from the relevant biological theory concerning individual organisms, i.e. *physiology*. Organisms are goal-directed in a sense that "they are disposed to adjust their behavior to environmental change in ways appropriate to a constant result, the goal" (Boorse 1977, p.555–556). This does not mean that the goals are necessarily conscious. It just (roughly) means – on my interpretation – that organisms predictably behave in ways that seem to target specific ends instead of behaving randomly.

There is a hierarchy of goals at different levels: cells are goal-directed to manufacturing certain compounds, thus contributing to higher-level goals like muscle contraction; these goals all contribute to overt behavior like web-spinning, nest-building or prey-catching. Overt behavior then contributes to the ultimate goals: survival and reproduction. Deciding what the ultimate goals at the apex of the hierarchy are is of great importance, since without knowing them, it would be impossible to decide what the functions of any organs are. Organs may have many different features and not all of them are functions. For example, a heart pumps blood and produces sounds. To decide which of these features heart's function is, one needs to know whether it is the blood pumping or the sound making which contributes to the *ultimate goal*. If one knows that the ultimate goal is survival, one may infer that the function of the heart is pumping blood, since pumping blood is consequentially related to survival while the sounds the heart makes are not. (Boorse 1977, p.556.)

Boorse notes that, to some extent, the highest-level goals are indeterminate and must be decided depending on biologist's interests. Most behavior of organisms usually contributes simultaneously to many goals, such as individual survival and reproduction, species survival, survival of the genes, ecological equilibrium and so forth. Different subfields of biology may use different goals as the focus of their function statements. But the subfield of physiology is the only one whose functions seem relevant to *individual health*[25]. Boorse asserts that physiology's functions are contributions to individual survival and reproduction. Hence, individual survival and reproduction are "chosen" as the ultimate goals of organisms in the BST[26]. Function statements will be value-free, since it is an empirical matter to find out what makes a causal contribution to a biological goal. In sum, functions of a trait are causal contributions it makes to organism's survival and reproduction. (Boorse 1977, p.556.) Then, *normal functions* are those functions which are working at a *statistically typical efficiency* (in a reference class). The statistical method used in assessing the normality of function is the aspect that makes the BST statistical.

For example, the function of an eye is seeing, since seeing contributes causally to survival or reproduction of the individual (Boorse 1977, p.557). An eye's *normal function* then, is seeing on the level that provides a statistically typical contribution to survival or reproduction in a reference class.

Boorse's view of biological functions plays a role when assessing the BST's possible value-ladenness. The criticisms related to functions and especially to the goals functions are assumed to contribute to, are addressed later in this essay.

4.3.2 Reference class

Function statements are about a trait's standard contribution (to a goal) in some population, e.g. species (Boorse 1977, p.556). A standard contribution is a contribution which is statistically typical in that population. By studying what these statistically typical contributions are, one can paint an idealized portrait of a typical member of a species. Biology textbooks provide these kinds of portraits which Boorse calls *species design*, i.e.:

[25] Saying that only physiology's functions seem relevant to health, seems to presume that it is known what health is. On the surface this seems to hint a possibility of a circular argument, as health and disease are the target concepts Boorse is trying to define using "function" as a part of the definition. Why Boorse's definition need not be circular, I will discuss in chapter 6.
[26] This has important consequences which are examined thoroughly in chapter 6.

> The typical hierarchy of interlocking functional systems that supports the life of organisms of that type. Each detail in this composite portrait is statistically normal within the species, though the portrait may not exactly resemble any species member. (Boorse 1977, p.557.)

Thus, what Boorse considers as species design, is kind of a theoretical specimen, which does not necessarily exist as such but represents the species norm in every aspect. This idea of species design is meant as an empirical ideal and Boorse suggests that it serves as the basis for health judgements in any species (Boorse 1977, p.557).

To make things a bit more complicated, there are some intraspecific differences which are significant enough to generate several different designs within a species, i.e. sex and age. There are certain female characteristics and certain male characteristics occurring consistently enough to be considered as constituting distinct functional designs. Human males typically have penises and testicles and females typically have ovaries and wombs and all these organs have their respective functions. Functional design also varies with age. For example, growing skeleton is a normal function for infants but not for adults. Adult males produce sperm while infants do not. Thus, it seems that species design is relative to age and sex. (Boorse 1977, p.558.)

This is the rationale behind Boorse's reference class, i.e. *an age group of a sex of a species*. Because people of different ages and different sexes have different functional designs, the judgements about what is normal and what is diseased should be made relative to sex and age. Otherwise absurd conclusions would have to be made, e.g., either adult males producing sperm or infants not producing sperm would have to be judged as diseased.

Boorse's "choice" of reference classes has generated criticism largely on the same grounds as his "choice" of goals. These criticisms are addressed later in this essay.

4.3.3 Health

Health in a reference class is *normal functional ability*, where normality is statistical, and functions are biological (Boorse 1977, p.542). When an organ is functioning at a statistically typical efficiency, i.e. making a statistically typical contribution to survival or reproduction of the individual in the reference class, it is functioning normally. If every part of the individual is functioning normally, the individual is *healthy*. This kind of complete *theoretical health* does not necessarily exist in nature, it is just an ideal. The fact that practically all individual organisms have *some disease* (Boorse 1977, p.547; 1997, p.85), i.e. subnormal functioning in at least one of their parts, is not a problem for the BST as long as all individuals (in a reference class) do not have the same disease (Boorse 1975, p.58).

That said, it is important to note that while statistical normality plays a major part in the BST, it fails as a necessary or a sufficient condition for health. It cannot be necessary, because unusual conditions such as type O blood or red hair may be perfectly healthy. It cannot be sufficient, because in certain populations unhealthy conditions may be typical. (Boorse 1977, p.546.) The fact that statistical normality cannot be applied as a necessary condition for health is not a problem for the BST, since the BST does not imply statistical normality as a necessary condition. According to the BST, any statistically abnormal feature may be perfectly healthy, *as long as it does not reduce any functions below typical efficiency*, i.e. does not decrease the contributions of any organs to the goals – survival or reproduction – below typical levels in the reference class. In the BST, statistical normality is necessary for health only in relation to functions as defined.

That statistical normality cannot be sufficient for health is a more complicated issue and might pose a problem for the BST. I already mentioned that the clause about "atypical environments" in the BST is meant to deal with those situations where a certain rare environmental factor causes pathology in such a high percentage of population (in that environment) as to make the pathological condition in question statistically normal. The clause states that in these unusual environments, statistically normal conditions may still be pathological. Thus, the clause enables the BST to classify conditions like heat exhaustion and mountain sickness as pathological, despite the fact that they are statistically nor-

mal reactions in the environments they occur. Hence, statistical normality cannot be sufficient for health by the BST. For criticism of Boorse's solution to the challenges posed by diverse environments and situations, see Kingma (2016).

One might still object that some conditions, such as high blood pressure or osteoporosis, might be statistically normal in some reference classes, irrespective of the environment. These kinds of "typical diseases" Boorse admits being a problem for the BST in its current form. Thus, Boorse has suggested a revision for the BST. He asserts that biologically and medically reasonable reference classes are such that "typical diseases" very rarely occur. When they do occur, most examples are diseases of old age. So, the problem would mostly disappear if the BST made young adults the standard for all adults. Then old people would not be considered as a separate reference class and diseases related to ageing would be diseases by the BST, even if they were statistically normal in a group of old people. (Boorse 2014, p.714.) I believe this kind of revision would at least partially resolve the problem posed by typical diseases. Whether it in fact does and how large a problem remains, is left for future research.

Boorse distinguishes his *theoretical health* concept, which is *absence of disease*, from *practical health*, which is roughly *absence of treatable illness* (Boorse 1977, p.542). That every individual might be somewhat diseased according to the BST, does not imply every individual should be given medical attention. For those kinds of normative treatment decisions Boorse recognizes separate "disease-plus" concepts, such as *illness*, *diagnostic normality* and *therapeutic normality* (Boorse 2014, p.684–685). These concepts are not of great importance for my purposes, as Boorse concedes they are value-laden. The relevant question is whether the concept of theoretical health defined above is also value-laden.

4.3.4 Disease

The concept of disease (or pathological condition) then follows from the above analysis of health[27]. If health is normal functioning, i.e. an organ making a statistically typical

[27] The standard way of looking at the relationship between health and disease is the other way around. This manifests in the phrase "health is the absence of disease" as Boorse notes (1977, p.542). But as health and disease are in Boorse's analysis mutually exclusive counterparts in the sense that an organ cannot be healthy and diseased at the same time and it must be one or the other, I don't see a problem in turning the direction of reasoning around. It must be kept in mind though, that deviating from normal functioning does not straightforwardly imply disease, as I explain in my analysis.

contribution to survival or reproduction of the individual in the reference class, then disease is *subnormal functioning*, i.e. an organ working below the statistically typical efficiency in a reference class in which the individual in question belongs to. Here it must be emphasized that the fact that "normal function" is assessed by statistical means does not mean one can just infer that anything outside the statistically typical range is always diseased or even a risk factor.

For example, let's look at two different groups of traits. First, traits like *body mass*, *blood pressure* and *resting heart rate* all have a statistical normal range, which for body mass is around BMI 19–25 (Iyengar et al. 2017; Gavin et al. 2015), for systolic blood pressure 90–130 mmHg (Xhignesse & Krzesinski 2018) and for resting heart rate 50–90 bpm (Nanchen 2018). Crossing either the upper or the lower bound might in all three cases be a sign of subnormal functioning, i.e. contributing to survival or reproduction below a typical level of efficiency, i.e. a disease. But, if we look at another group of traits, *intelligence*, *muscle strength* and *eyesight*, the analysis looks different. All traits in this latter group certainly have statistically typical "normal ranges" (whatever the exact ranges are) as well as the traits in the first group. But only crossing the *lower bound* of any of these traits' normal ranges is considered a sign of disease (Boorse 1977, p.559). If one is of low enough intelligence, one is considered *mentally retarded* (Shandera et al. 2010). Low muscle strength may be a sign of *myotonic muscular dystrophy* (Kataoka & Ueno 2017). If one does not see objects far enough or near enough, one might have some of the many possible *vision disorders*. But nobody considers *uncommonly high intellect* or *strength* or *uncommonly sharp vision* – either near or far – as a sign of any kind of disease.

The point of the above example is that not any deviation from normal functioning is a disease in a statistical model. Diseases are those deviations from normal functioning, which lead to lesser than typical contribution to the ultimate goals of the organism, i.e. survival and reproduction. The normal statistical range of any trait does not by itself provide the information needed to determine whether crossing the upper bound and/or the lower bound implies a disease.

4.4 The difference between disease and illness

While the BST's concepts of theoretical health and disease are my main focus here, I will briefly explain how Boorse distinguishes the latter from the more practical concept of

illness. This distinction highlights the reasons why Boorse thinks that some of the normativists – the antipsychiatrists for example – got their ideas wrong. This may also clear some misunderstanding rising out from the fact that medical language many times *sounds normative*. Because treatment decisions are normative, it is understandable that there is an illness-concept which is tied to these normative decisions and this concept is used in some medical and lay contexts. But the point is that this fact does not rule out the possibility of a theoretical and value-free concept of disease, which in principle could apply in both physical and mental areas of medicine.

In Boorse's view, the theoretical concept of disease is *value-free* as it allows (in principle) an organism's disease status to be settled by the methods of natural science. He argues that the opposite view – e.g. that of the normativists and some antipsychiatrists – is due to a failure to distinguish between this purely theoretical concept of *disease* and a much narrower idea of an *illness.* (Boorse 1976, p.63.) According to Boorse, diseases – which might be quite harmless minor deviations from the norm which everyone has in some degree – become illnesses only when they satisfy certain normative conditions:

> A *disease* is an ***illness*** only if it is serious enough to be incapacitating, and therefore is
>
> (i) undesirable for its bearer;
>
> (ii) a title to special treatment; and
>
> (iii) a valid excuse for normally criticizable behavior. (Boorse 1975, p.61.)

Thus, illnesses are a narrower subgroup of disease. While diseases are more theoretical entities with no necessary practical or social implications, illnesses are the ones that usually mean something to lay people and are of significance to functions of social institutions. The criteria Boorse lays out makes it explicit that (i) to call a biologically disadvantageous deviation from normality (i.e. disease) an illness, is to suppose there is an element of harm present, so that the deviation in question becomes undesirable for the individual concerned[28]. The second criterion (ii) makes the case that *illnesses* – not *diseases* – are conditions which are serious enough to warrant special treatment. The third

[28] It is essential here that the undesirability is assessed from the individual's point of view (and not society's).

criterion (iii) seems quite a strong ethical presumption about the relationship between illness and responsibility and Boorse makes its controversiality clear (1975, p.66). It should be emphasized that these conditions are part of Boorse's *analysis of disease and illness concepts*, not his own evaluative judgements. Thus, for example, in (iii) Boorse is not arguing that people with illnesses *should* be morally excused when they act badly. He is just claiming that in our discourse *illness* is commonly used in such a way that it is presumed to entail (at least the possibility of) diminished responsibility.

When constructing his model of disease Boorse was motivated by the antipsychiatrist claims about the value-ladenness of psychiatry's illness concept. Boorse's purpose was to show that *health* and *disease* can be theoretically defined as value-free (Boorse 1975) and that "the functional idea of health in physiological medicine applies as straightforwardly to the mind as to the body" (Boorse 1976, p.62) even if "mental illness" may often be value-laden in practical contexts.

Despite his position that functionally defined theoretical health applies smoothly to mental domain, Boorse is quite pessimistic about applying the analogy between the physical and the mental to *illness concepts* defined the way above. The reason is that the relation between persons and their illnesses is conceived on the model of their relation to their bodies. This relation functions as the base for justifying the logic behind conditions (i), (ii) and (iii) defining illness in the physical domain. But as the relation of persons to their bodies differs in important ways from their relation to their minds, the same logic may not apply in the mental domain. (Boorse 1975, p.62.)

A couple of examples may clarify why the analogy is problematic:

a) The condition (i) states that to call a disease an illness, it should be *undesirable for its bearer*. In the physical domain there is not a lot of controversy whether diseases are undesirable[29] and thus this condition is mostly unproblematic. But what about "mental illnesses" like sexual deviances? If I have a "fetishistic disorder", e.g. "foot-fetish" (still listed as a mental disorder in DSM-5 (2013)) is it undesirable for me? Is it undesirable *for me* to desire (and get satisfaction of) something which is deviant in my society? It might be or might not (depending

[29] Not to say there aren't physical diseases which may be desirable. Boorse mentions "alcoholic intoxication" (1975, p.61). It is a disease satisfying Boorse's technical definition, but whether it is desirable and should or should not be called an illness depends on the context.

on overall consequences), but there is a risk of confusing things which are undesirable for the society with the things that are undesirable for the individual. It is not obvious why we should *desire* our desires to conform to the species typical (Boorse 1975, p.63) or even society typical norm.

b) The condition (ii) states that to call a disease an illness, it should act as *a title to special treatment*. The emphasis here is on "special", limiting the treatment in question for those *seriously ill*. The problem is that some mental illnesses, such as "neurosis" and "paranoid personality", have been considered *statistically normal* in some societies. As 1) statistically normal conditions can be diseases (by the BST) only if they are due to environmental factors and 2) whole societies may be considered as disease generating environments affecting most of the population 3) then if the disease would be serious enough to warrant treatment and to be called an illness, we would have to call everybody in that society ill and give treatment for everybody. (Boorse 1975, p.65.) Then the treatment would not be "special" (limiting kind), it would be "standard" for all[30].

c) The condition (iii) states that an illness can be "*a valid excuse for normally criticizable behavior*". So, if I have lost my legs in a car crash, most people would excuse me of *not running* to help a drowning person (and there is a clear rationale). But if I am a *psychopath* i.e. I have a diagnosis for "antisocial personality disorder" (DSM-5 2013) and I happen to like watching people drowning while eating popcorn, it is not at all clear many people would excuse my behavior (or if they surprisingly would, what would be their rationale?).

Hence, it is good to keep in mind that while Boorse is a strong proponent of the possibility of value-free health and disease, he is not claiming there are no problems in the psychiatric classifications or in the way *mental illness* is sometimes interpreted analogically to *physical disease*.

To sum it up, in Boorse's view there is a distinction between *theoretical health* and *practical health*; *disease* refers to the absence of the former and *illness* to the absence of the

[30] This argument may be outdated due to the developments in psychiatry. "Neurosis" was not a part of the official nomenclature since DSM-III (1980). More research needed.

latter (Boorse 1975, p.56). This is where the antipsychiatrists go wrong; they do not distinguish between these theoretical and practical meanings of health and therefore their analysis of disease and/or illness is mistaken (Boorse 1975, p.49). Boorse acknowledges that in clinical and other practical contexts there might certainly be evaluative judgements at play when deciding who is ill and who is not, and these judgements might be problematic. The common use of the word "illness" and its consequences are thus open to antipsychiatrist (or normativist) criticism. But contrary to antipsychiatrist claims, Boorse argues there is nothing ruling out the possibility of a theoretical non-evaluative description of health and disease. (Boorse 1975, p.55–56.)

Now that I have provided an overview of the BST and Boorse's distinction between the value-free theoretical concept of disease and the value-laden practical concept of illness, it is time to take a deeper look into the arguments for and against the BST's value-ladenness. These arguments are presented and analyzed in the following chapters.

5 Varieties of value-ladenness

The liveliest discussion in recent decades concerning the concept of disease has been revolving between the naturalists and their opponents. As noted, naturalists claim that health and disease are value-free concepts that pick out a natural, real-world category (Kingma 2017, p.49). The opposing side consists of people criticizing the current naturalist accounts and saying they fail to be value-free in one way or the other. Some of these critics are explicit normativists (e.g. Cooper, Fulford and Reznek) and some are undecided (e.g. Kingma). Boorse positions himself strictly to the naturalist camp. As explained earlier in this essay, his influential work in the 1970s (Boorse 1975; 1976a; 1976b & 1977) was originally motivated by the antipsychiatrist claims about the value-ladenness of mental illness; with the Biostatistical Model, Boorse wanted to show how theoretical *disease* could be defined value-free, even if *illness* is often intertwined with values in practical contexts (Fulford 2001, p.80–81). Boorse claimed that the value-laden use of illness-concept in common language does not necessarily imply that medicine's theoretical disease-concept also is value-laden.

The BST has generated lots of criticism after the model was introduced in the 1970s. Broadly, the criticism falls into four categories (Kingma 2017, p.53):

1. Technical objections
2. Complaints about value-ladenness
3. Charges of employing bad biology
4. Specific counterexamples

Recent debate in philosophy of medicine and psychiatry has focused on the first two categories. Technical objections are mostly about (i) where to locate the distinction between health and disease, (ii) whether the BST can accommodate environmental variation and (iii) how the model handles "ubiquitous diseases", i.e. situations where a seemingly diseased condition is statistically normal. (Kingma 2017, p.53–54.) These kinds of objections may have significance for the validity and practical usefulness of the BST, but for time/space-resource related reasons, they must be addressed elsewhere[31]. The questions concerning value-ladenness, which are my main interest here, can mostly be addressed separately from the criticized technical details. For the details of technical objections see Nordenfelt (1993), Schwartz (2007), Guerrero (2010) and Kingma (2010 & 2016) and for replies to those and all kinds of criticisms see Boorse's "Rebuttals" (1997 & 2014).

The third category – bad biology – is mostly about what is the correct account of biological function. There is a vast literature about function statements (see Garson 2016) and it mostly does not touch the issues connected to value-ladenness. Boorse asserts that the majority view in function literature is that biological function statements are value-free (Boorse 1997, p.58). Should function statements turn out normative, then the BST would imply health and disease are normative too and the BST would collapse. For better or worse, I'll leave the question about the correct account of function to biologists and philosophers of biology and settle here for the presumption that functions can be defined in value-free terms.

The fourth category – specific counterexamples – is interesting but too diverse to be properly addressed in this essay. Besides, some of those examples come up in my analysis regarding the value-ladenness objection.

[31] A brief look into these technical issues is provided in chapter 4 of this thesis.

My focus is on the second type of criticism, concerning the various claims about the value-ladenness of Boorse's disease concept. These claims may be divided into three categories: a) *claims about value-laden choices* b) *claims about value-laden use of language* and c) *the general claims about the value-ladenness of science*.

In this chapter, I will start from the last one and address it only briefly, arguing that the possibility of natural sciences being generally value-laden is inconsequential in the context of the current debate about the BST. Then, I will turn to the question about the significance of the way we use language and propose that evaluative sounding language is not necessarily a problem for the BST, *if* the language judged as evaluative can be replaced with descriptive language *without loss of meaning*.

In chapter 6, I will give a thorough assessment of the claims about the BST's value-ladenness in the choice-level. Assessing these claims is a difficult and complex task, because it involves lots of interpretation of a conversation where people are talking somewhat past each other. On the other hand, I think it is also the most important of all the tasks concerning the criticisms about value-ladenness, as it touches the questions about what it means for health and disease to be judged as value-laden and why it should matter if they are. As Kincaid (2007, p.228–229) notes, even if values are involved in a certain scientific context, it is unclear *what role they play if any*. The *meaning* and *significance* of value-ladenness in the context of health concepts are the puzzles I see most value in solving.

5.1 All science as fundamentally value-laden

Kincaid suggests that the majority view in the literature of philosophy of science is that science is essentially value-laden (Kincaid et al. 2007, p.219). There are always at least some epistemic values and interests in the motivational-level at play in scientific processes (Doppelt in Kincaid et al. 2007, p.190). One may decide to call "value-laden" all the products of these scientific processes driven by values and interests. On these grounds, one may infer that the BST must be value-laden, since the concepts it defines are products of a value-laden scientific process described above.

In the context of discussion about the BST, Agich (1983) and Stempsey (2000) hold this kind of view, just to name a few. To give a taste, Stempsey claims that "our very descriptions of anatomy and physiology are themselves value-laden" (Stempsey 2000, p.327).

Hence, were Boorse to try defining "disease" based on physiological theory, his end product would be doomed value-laden from the beginning.

Boorse does not have sympathy for this kind of view. He notes that the BST is value-laden insofar as biological function statements are value-laden. But in the function literature, holding them to be so is a minority position. (Boorse 1997, p.58.) Importantly, Boorse is only trying to argue that health and disease are not *especially* value-laden in relation to scientific concepts in general (Boorse 2014, p.713). Thus, if Agich, Stempsey or others are only claiming that health and disease – and hence the BST – are value-laden because they involve scientific concepts which themselves are fundamentally value-laden, Boorse thinks the claim does not pose a problem for the BST. If the BST is only as value-laden as sciences such as physics or chemistry, it achieves "everything he ever dreamt of for it"[32] (Boorse 1997, p.56).

I do not think the criticism resting on value-ladenness of all science deserves a deeper examination here. Whether natural sciences are value-laden in any relevant ways may be an interesting question in itself. But from the start, the controversies motivating Boorse's work have clearly been revolving around the possibility of health concepts' being *especially normative*. The contentious question has been whether health concepts are value-laden in a certain special way that may lead to important sociopolitical implications. Claiming that all science is value-laden because of the possibility of interests and epistemic values driving research says not much about the question at hand. The possible value-ladenness of health concepts must be assessed by critically examining how they are constructed and defined taking the context into account.

5.2 Evaluative language and descriptivism

The second type of criticism to be addressed here is linguistic-analytic and it focuses on how language is used and how that usage relates to definitions of concepts. In the context of defining health concepts, this line of thought is elegantly represented by K.W.M Fulford (1989 & 2001). Before getting into the specifics, I'll present Fulford's general argument against the BST which consists of three inter-connected points:

[32] Of course, it could be argued that biology is more value-laden than physics or chemistry and therefore grounding the BST in biology makes it more value-laden than "real" natural science in general, but I cannot address this kind of question here.

> 1) A naturalist (value-excluding) medical model of disease leads into inconsistencies.
>
> 2) The only self-consistent form of the medical model, a descriptivist (value-entailing) medical model, fails when applied to the problems raised by mental disorder.
>
> 3) The "part function" account of dysfunction on which Boorse relies for his naturalist (value-excluding) medical model, fails when applied to the problems raised by the central kind of mental disorder, psychosis. (Fulford 2001, p.80.)

As we can see, Fulford thinks that the BST is inconsistent, that the only possible self-consistent model is descriptivist, would entail values and fails in the context of mental disorder, and that the BST which relies on part-functions does not apply to one central mental disorder, namely psychosis. The logic of this argument opens up only after I will (shortly) explain what Fulford means by "value-excluding naturalism" and "value-entailing descriptivism". The point here is that if this criticism hits its target, it is quite damning for the BST, recalling that Boorse cannot accept values in his model and would like to apply it in the domain of psychiatry (Boorse 1997, p.13).

The critical problem Fulford sees in Boorse's work on health concepts is the inconsistent use of language. While Boorse is careful to define "disease" in value-free terms, he nonetheless uses *disease* with clear evaluative connotations. The example Fulford gives, concerns the way how Boorse first defines disease value-free as a "deviation from functional norms" and three lines later calls disease "deficiency in functional efficiency". Another example is Boorse writing of disease being due to "environmental causes" which after few lines becomes "hostile environment"[33]. (Fulford 2001, p.81.)

Fulford interprets "deficiency", "efficiency" and "hostile" as value-laden terms. From his linguistic-analytic perspective, these kinds of shifts from value-free definitions to value-laden uses of language are highly significant. As people are better at using "high-level concepts" than at defining them, the words people use are surer guide to concepts' meanings than attempts to define them. Hence, using "disease" with evaluative connotations

[33] Sadler & Agich (1995) see the same problem in Wakefield's harmful dysfunction -analysis of disorder. For Wakefield's answer which resembles Boorse's, see Wakefield (1995).

may signal that there is an essential evaluative element in the meaning of the term. (Fulford 2001, p.81.)

Boorse does not see his use of language as problematic, because – as he claims – where his terms slip to the normative side, they can be easily replaced with descriptive terms. He claims the evaluative rhetoric can be eliminated without a loss. (Boorse 1997, p.20.) By "loss", Boorse refers to "loss of meaning". If that kind of elimination could be done, there would not be a problem in the sense of Fulford's criticism as there wasn't any suspicious evaluative terms left. But Fulford does not accept this kind of elimination as a solution.

Fulford explains that in principle Boorse could use a descriptive strategy to evade the problems connected to value-ladenness of his model. The relevant form of descriptivism that would offer Boorse a "way out", works in the following way: it defines the value terms in question by reference to the descriptive criteria for the value judgements they express (Fulford 2001, p.81). This sounds quite a lot like the solution by elimination Boorse offered. But there is an important difference.

The critical distinction Fulford makes is between *value-excluding naturalism* and *value-entailing descriptivism* (Fulford 2001, p.82) and Fulford thinks only the latter would save the BST. When Boorse says that he can eliminate evaluative sounding terms easily and replace them with clearly nonevaluative, descriptive ones without loss of meaning (Boorse 1997, p.20), this is what Fulford calls *value-excluding naturalism*, not *descriptivism*. This position is naturalist in a sense that it assumes that the possibility of replacing the evaluative sounding terms with descriptive ones (without loss of meaning) shows that the meaning of the originally used terms in that context is value-free. But from Fulford's linguistic-analytic perspective, the true guide to terms' meanings is how they are *used*, not how they are *defined*. So, in contrast to naturalist position, Fulford argues that the fact that a term can be redefined in value-free terms does not show the term itself is value-free. The true meaning of a term is revealed by its use.

The "way out" Fulford offers for Boorse is the *value-entailing descriptivism* mentioned above. The idea of value-entailing descriptivism is that under certain circumstances, a given set of facts *implies* a given value judgement. The evaluative meaning of the word is captured in descriptive criteria for the value judgement the word in question is used to express. If this kind of descriptive strategy is accepted, there is no conflict between the

value-free description and the evaluative use of the term. As Fulford sees it, this value-entailing descriptive strategy would give Boorse a value-free science of health at the same time it allowed him to continue using health vocabulary with evaluative connotations. There would not be inconsistencies, because the descriptive terms used in the value-free definitions would imply the same values that Boorse's evaluative use of the terms described imply. (Fulford 2001, p.82.)

To illustrate this point, Fulford uses IBM's (a tech company) share price as an example. The evaluative-sounding judgement that the share price is "better by 5/8" can be replaced by its descriptive criteria "higher by 5/8", which describes the share price as having gone up by 5/8. The description itself, "higher by 5/8", is value-free as it only states an arithmetic fact. In a case like this, a naturalist (such as Boorse) would say that the use of the term "better" in this context is just "harmless rhetoric", as it can be replaced with a descriptive value-free term "higher". In contrast, a (value-entailing) descriptivist would say that "higher by 5/8" is descriptive indeed, but at the same time it is the descriptive criterion for the value-judgement "better by 5/8". There is no loss of meaning in the process, as "higher by 5/8" implies "better by 5/8" in this context. (Fulford 2001, p.82.) Hence, there is no inconsistency as the definition and the use of the term entail the same (evaluative) meaning.

Fulford brings up an additional important point: the key condition required for the descriptive strategy to work is that the values redefined in descriptive terms are *shared*. He argues that this condition is usually satisfied in physical medicine but not in psychiatry. Diseases like cancer or heart attacks are considered bad by anyone. The same does not apply to mental conditions, which are about desires, beliefs, emotions, motivations and so forth. Therefore, the (value-entailing) descriptive strategy is not available in the context of psychiatry. (Fulford 2001, p.82.)

I am not going to assess the specific problems related to mental disorders here, but I'll briefly comment Fulford's point about shared values. First, his example[34] using IBM's share price is in fact accidentally a good example for showing how the assumption that values are shared may *fail* and how that failure may cause problems. The reason why "higher by 5/8" implies "better by 5/8" is – according to Fulford – that the descriptive

[34] Which apparently is originally coined by Boorse (Fulford 2001, p.82).

criteria for "better" in the context of share prices is fixed by conventional definition (Fulford 2001, p.82). I interpret Fulford meaning by "conventional definition" that since everybody agrees higher is better in the context of stock prices, "higher" and "better" are interchangeable in that context. The value of "higher price" is shared in the context of stock markets, thinks Fulford.

But is it? If you are *short* on IBM, meaning you have made a bet *against* IBM's share price, then you are hoping for the share to decrease in value, i.e. the price to go *lower*. I do not think short-sellers would agree that if "better" is replaced with "higher", there is no loss of meaning. There are probably many other people too who would like to see a certain share to go lower instead of higher (potential buyers, competitors, some socialists etc.), so I would say there is no universally shared intuition that higher is better in the context of stock markets.

The problem in Fulford's example is that the value that *is* shared behind the judgement of "better" in the example is fundamentally the value of more money (consumption opportunities) over less money. But whether the suggested descriptive criteria of "better" (i.e. "higher") implies more money and thus more consumption opportunities, depends on one's position in the market. If one's position is *long*, then higher implies better. But if one's position is *short*, then higher implies *worse*. That is why the example fails to reflect a context in which the presented values are truly shared and the descriptive criteria for the value judgement undisputed.

Even if the specific example was bad, it managed to illustrate Fulford's point. To be able to give a value-free description of an evaluative term without loss of meaning, the values implied by the term should be shared. Otherwise there would be no agreement of the terms in which the evaluative term should be value-freely described. I can accept this point. But what is important is that *Boorse can probably accept it too*. It seems to me that Fulford presupposes a major premise that Boorse is disputing.

Recall that the problem Fulford sees in Boorse's account of disease is that it is *inconsistent*. And by inconsistent, Fulford means that while Boorse has shown that he can give value-free descriptions of certain relevant terms, he continues using them with evaluative connotations. And this tension between descriptive definitions and the evaluative usage

of terms is significant from linguistic-analytic perspective. According to Fulford, the account would be consistent only if Boorse would adopt a value-entailing descriptive stance instead of his value-excluding naturalist stance.

The critical problem in Fulford's argument is that his value-entailing descriptivism is an analytic strategy *meant to deal with value terms* by giving them factual descriptions which imply the values in question. Fulford seems to *presume* that the terms Boorse uses in his analysis of health concepts – like efficiency or hostile – *are value terms*. Then he logically goes on explaining why Boorse needs to adopt value-entailing descriptivism to remain consistent between his definitions and use of language. Value-excluding naturalism won't do, because it would not capture the meaning of the value term that is described.

As sophisticated as Fulford's analysis is on the surface, I cannot help but noting that it seems to *presume* the normativism and the value-ladenness of health vocabulary which is what Boorse explicitly *denies*. As Boorse himself puts it:

> (Fulford) views normativism about health as nearly axiomatic. At the bottom, he cannot grasp how anyone could fail to be a normativist, and he thinks I am one in spite of myself (Boorse 1997, p.22).

Even if little harsh, my impression is that Boorse does not grossly misrepresent Fulford's case here.

Boorse's claim is that when he uses those terms which Fulford sees as evaluative (deficiency, hostile etc.), he is *not* using them with *any* "evaluative connotations". He is using language in a way he thinks is more convenient. By "deficiency" of function – for example – he just means "less function, less contribution to the goals" (Boorse 1997, p.21). That is just an arithmetic concept, not evaluative. At least in the context and in the way Boorse uses it. As Boorse sees it, he could as well say "less function" instead of "deficiency of function" without any loss of meaning. That is why he calls himself a naturalist and not a descriptivist. The latter usually refers to the philosophical position according to which *value terms can be descriptively defined* (Boorse 1997, p.21). In contrast, Boorse's position is that *no value term can be descriptively defined* (Boorse 1997, p.22). Hence, it is clear that Boorse's project is not redefining value terms descriptively. He is trying to show that in medical theory, health, disease and other relevant terms *are not value terms*.

Consequently, I think Fulford's criticism would be relevant only if one accepted as a *premise* that terms such as "deficiency" and "hostile" are value terms in the context they are used. If they are value terms in that context, then the only way to describe them in value-free terms without loss of meaning is to adopt Fulford's perspective of value-entailing descriptivism. But Boorse *does not accept* that "deficiency" and "hostile" etc. are value-terms in the context he uses them. They are just a rhetorical option he has chosen to sound less clumsy. Nothing is lost if they are replaced with clearly descriptive terms. Hence, Boorse is a naturalist, not a descriptivist, and his model is self-consistent as it defines *value-free concepts in value-free terms*. No evaluative meaning is lost on the way.

6 Goals and reference classes

6.1 The choice of goals

Commonly, the claims about the BST's value-ladenness seize on the fact that there is – or at least seems to be – a choice involved in determining the goals or the reference classes (Kingma 2017, p.53): in the BST, the fundamental goals are stated to be survival and reproduction and the reference classes are constructed as age-groups of a sex of a species (Boorse 2014, p.684). The goals and the reference classes could have been chosen otherwise and the actual choices made involve value judgements, the critics claim. Furthermore, had the choices been different, different groups of people could have been labeled as "healthy" and "diseased". Thus, what is considered as diseased according to the BST, depends partly on the value-laden choice-process behind the goals and the reference classes the model adopts. Hence – the critics say – Boorse's concept of disease is not value-free as it is supposed to be.

I will first deal with the choice of goals and then with the choice of reference classes.

The two most prominent recent criticisms concerned with the BST's choice of goals are provided by Devito (2000) and Ereshefsky (2009). Ereshefsky argues that the BST assumes biological fitness, i.e. survival and reproduction, as *the* goal of human and all life. But, he notes, humans might have multiple goals and some of them may run counter to individual fitness. Humans may even sacrifice their reproductive ability for other pursuits. (Ereshefky 2009, p.223.) For example, one might get a vasectomy to make sure he does not have more offspring in order to enjoy more leisure instead. Or one might choose a

spiritual life in celibacy in a monastery. It is not obvious why survival and reproduction should be chosen over other possible goals; aligning health with biological fitness is not dictated by biological theory (Ereshefsky 2009, p.225).

According to Ereshefsky, Boorse's response here is that as "fitness decreasing goals" (e.g. some social or well-being related goals) are outside the realm of biology, they should not be considered relevant for medicine. Ereshefsky objects to this line of response, pointing to WHO's (World Health Organization) definitions of "health" and "disease" which cite "physical, mental, and *social well-being*". If WHO's definitions are part of scientific literature on medicine, shouldn't goals implied by those definitions (e.g. social goals too) be considered relevant for medicine, asks Ereshefsky. (Ereshefsky 2009, p.223.)

Ereshefsky also objects on a more general level to Boorse's claim that biological fitness is *the* biological goal of humans and all organisms. Biologists describe organisms as having many types of states and many of those states have nothing to do with fitness. Thus, to say that the goals of an organism are survival and reproduction – instead of other kinds of biological states – is to make a choice between options. And the justification for that choice comes from outside of biology. As naturalism demands that only science – in this case biology – may tell us what "health" is, making the choice on other than scientific grounds shows how Boorse's account fails to be naturalistic. (Ereshefsky 2009, p.223.)

DeVito's criticism takes the same form as Ereshefsky's general criticism above but is presented a bit more abstractly. DeVito makes the point that:

> When there are multiple sets of criteria that adequately divide the world into things to which the concept applies and things to which it does not, we can choose one of these concepts as the ´correct´ or accepted concept. Values can enter at this choice-level. (DeVito 2000, p.541–542.)

The interpretation in this context being that if biology offers multiple ways of setting the criteria for health, we can select one specific way and accept the concept derived from it. But in doing that, values are involved as biological theory does not entail instructions for making the choice.

I see problems in both kinds of objections described above. The first one by Ereshefsky, resting on WHO's definition of health, gives the definition too much credit in determining the area of interest of theoretical medicine. The second and more general one by both

Ereshefsky and Devito is complex and difficult to tackle. The difficulties in answering to this objection are partly due to the general complexities of the problem and partly due to the fact that there seems to be some misinterpretations and talking past each other from both sides of the argument.

Next, I will try to clarify the conversation and shed some light on the sources of disagreement. I will also outline some potential arguments for the BST's current "choice" of goals. Importantly, I will try to assess what it means to be value-laden in a way the critics accuse the BST of and whether there is a good enough rationale for the BST's choice for the choice to be judged as trivial or at least not as damning as the critics seem to think. Before getting into these issues, I will explain why I do not think WHO's definition of health has much weight in the on-going discussion.

6.1.1 WHO´s definition of health

As WHO is an influential organization promoting global public health, there does not seem to be *prima facie* reasons to dismiss its definition of health in the context of the conversation about what health and disease are. But on a closer examination, there might be good reasons to do that after all. Let us first look at the WHO's Preamble to the Constitution from 1948, including the part which became to be taken as WHO's definition of health:

> Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.
> The enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being without distinction of race, religion, political belief, economic or social condition.
> The health of all peoples is fundamental to the attainment of peace and security and is dependent upon the fullest co-operation of individuals and States. The achievement of any State in the promotion and protection of health is of value to all. (WHO 1948.)

The first clause, health as a "state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" was picked up by philosophers as a

(WHO's) definition of health and started the decades long conversation around the concepts of health and disease partially covered in this thesis. But as Bickenbach (2017, p.962) notes, WHO's definition has been considered flawed and as such only taken as a starting point for the philosophical analysis, highlighting the conceptual issues in a need of clarification. Nordenfelt (2017, p.34) adds that WHO's definition "has been much criticized and has had rather little clinical significance". Even Wakefield himself – who is the one Ereshefsky (2009, p.223) refers to when making his WHO-related argument – presents WHO's definition in quite a critical tone, noting that "it appears to assume that disorder is any deviation from a completely desirable and ideal state" (Wakefield 1992, p.376).

As that kind of concept of disorder would imply that almost everybody is disordered, I interpret Wakefield's comment as mocking WHO's definition[35]. Generally, I think it is safe to say that in no way was the definition ever thought as a correct description of medicine's usage of *health* and thus it is questionable to give much weight to it when assessing medicine's area of interest.

The obvious sources of criticism in the definition are its breadth ("physical, mental and social well-being") and its unrealistically high threshold ("complete") of good health (Bickenbach 2017, p.962; Boorse 1975, p.60). Including "social well-being" in the definition of health seems to open the door to strange interpretations of what is "diseased". For example, one might be reasonably considered as not being "socially well", if one is poor or oppressed. But it would be very unorthodox use of language to call the poor or the oppressed "diseased". It seems obvious that medical theorists or physicians do not mean poor or oppressed people when they speak of people suffering from disease (despite those groups probably being somewhat correlated). The definition including "social well-being" as a criterion for health blurs the line between social and medical issues (Callahan 1973) and describes neither the actual medical usage of health-concepts nor any optional usage preferred by acquainted philosophers.

Secondly, according to WHO's definition, to be considered healthy seems to demand one to achieve "complete" – perhaps "perfect" or close to it – well-being in physical, mental

[35] This seems especially obvious considering Wakefield's major contributions to the literature as a stark opponent of diagnostic proliferation (see Wakefield 2010; 2011 & 2016). Wakefield's influential Harmful Dysfunction Analysis of disease is motivated by the goal of separating medical problems from social issues (see Wakefield 1992 & 2007).

and social domains. One is tempted to doubt if there is such a thing as "complete well-being" and specifically whether any human fits into this description. At most, it qualifies only a tiny minority of population as healthy, which would be in a stark contrast to our common usage of health vocabulary[36].

Thirdly, Boorse has in his recent works emphasized that his health and disease concepts target *pathologists' usage* of those terms (Boorse 1997, p.17). In fact, Boorse does not even use *disease* anymore in the latest forms of the BST (Boorse 2014, p.684). He replaced *disease* with *pathological condition* as he decided the broad usage of *disease* to be atypical of medical writing[37] (Boorse 2011, p.26). So even if WHO's definition would imply the possibility of social goals being relevant to medicine's overall area of interest, it would be unclear whether social goals would have any relevance from the more limited pathological perspective. Hence, in a minimum, Ereshefsky would have to adjust his criticism a little, as to specifically target the choice of this pathological perspective, which might or might not in itself entail the justification for seeing survival and reproduction as *the* goals of the organism.

In sum, from the fact that WHO is a prestigious organization involved in the areas of health and medicine, it does not automatically follow that its old informal definition of health should have much weight when trying to assess what is included in the current theoretical domain of medicine. Specifically, the fact that the definition includes "social well-being" as a criterion for health does not show that modern medical theory concerns or should concern itself with social goals, since the criterion in question is perhaps the most criticized detail of the whole vastly criticized definition. Finally, even if WHO's definition (against the odds) turned out to imply that theoretical medicine has some interest in social well-being, that would not necessarily imply much for Boorsian *pathologist's perspective*. Hence, I do not see Ereshefsky's criticism based on WHO's definition of health posing any serious threat to the BST.

[36] Whether conflicting with common language is considered a problem, depends on one's stance towards conservatism vs. revisionism debate, discussed in chapter 3 of this thesis (see also Murphy 2015, section 2. "Naturalism and Constructivism").

[37] This must not be taken as a change in the intended meaning of the concept though. The point of the change was only to be better aligned with common medical vocabulary.

6.1.2 The role of values in the choice-level

The second and the more important objection – presented by both Ereshefsky (2009) and Devito (2000)– is that there are many kinds of biological states of the organism biologists describe, and to choose survival and reproduction as goals from all those possible states is to make a value-laden choice. As there is nothing in biological theory itself implying what should be chosen, Boorse's theory fails to be naturalistic, as naturalism demands that biology itself should tell us what the goals are – and thus what *health* is.

So, this argument accepts as a premise that – in principle – it is justified for a naturalist to derive the goals from biological theory if possible. It just denies that the two particular goals – survival and reproduction – could be derived straightforwardly from the theory without evaluative considerations. The critics are saying that biology offers more than one way to conceptualize the facts and the way we choose determines what we end up considering the goals of the organism.

The objection rises three separate questions.

(1) First, is there really a *choice* involved or is it possible that the goals can be derived straightforwardly from biological theory?
(2) Second, if there is a choice, is it necessarily value-laden? Perhaps there could be choices not having anything to do with values.
(3) Third, if there is a value-laden choice, what is the *role* of values in it? Saying only that something is value-laden says very little, if it is not specified what kind of value-ladenness it is in question and what it implies.

To get some insight into these questions, it is probably good to start from Boorse's own comments on them. The natural places to look for the details of Boorse's defense are his two "Rebuttals on Health" (1997 & 2014), in which he answers to essentially all his critics after the original 1970s works until 2014. In the 1997 paper, he replies to roughly the same objections presented above about the choice of goals, earlier put forth by different writers (Engelhardt 1976; Brown 1985). In the second Rebuttal (2014), Boorse replies directly to DeVito and Ereshefsky, adding just a little to his arguments from 1997. Boorse's explanation on why the BST "chooses" survival and reproduction as goals touches in some way all the questions I presented above. So, let's get into the details of the explanation and Boorse's reply to the objection.

No choice at all

Boorse's first and most simple argument against the objection resting on choice of goals is that *in constructing the BST, **he** does not make a choice at all*. The BST's concept of disease is just the one that is best reconstructed from medical classifications. If there is a normative choice, it is medicine's original choice to combat disease instead of promoting it or striving for some other end. Boorse concedes that *medicine's choice to combat disease* might be a normative choice, but as such it does not show that the *meaning of disease* rests on a normative choice, unless one assumes the meaning to be fixed by medical practice. (Boorse 1997, p.25.) To make that assumption would be a mistake, as doctors have a notion of disease that is conceptually independent of medical treatment (Boorse 1997, p.26).

What Boorse is saying is that there in fact might be a choice involved, but if there is, it is not *his* – it is medicine's, and that choice does not imply the *concept of disease* to be value-laden. Medicine has "chosen" to apply physiology's goals (survival and reproduction) in its pursuit of combatting disease. Functions found in human physiology seem to serve the goals of survival and reproduction. That is why function statements in physiology and medicine can be explained as relative to these individual goals and not as relative to some other possible goals, e.g. survival of the species or some ecological goals. (Boorse 1997, p.28.) That is some reason to pick survival and reproduction instead of some other goals. It does not show the "choice" is value-free but at least gives an explanation why it was made. In addition, it is unclear if there are any reasonable options for physiology as medicine's background theory, *given medicine's interests in treating physiological problems*. If there is not, in what sense picking physiology was a value-laden choice? I will get back into this question shortly.

It seems sometimes quite difficult to get a good sense of what Boorse and his critics are exactly disagreeing about. I think this might be due to some confusion on the part of the critics about what Boorse is trying to do with the BST. As Boorse has often explicitly stated, his goal with the BST is to *analyze the concept of disease as it is used in medical theory*. He states this clearly already in his basic health paper (Boorse 1977, p.550–551) and reminds of it later, also quite recently (Boorse 2014, p.683 & p.693). What Boorse claims he is trying to do, is to work out the exact meaning (i.e. "analyze") of *disease* as

used in medicine and specifically by pathologists. This is not the same as trying to construct the best disease-concept one could imagine from the scratch. Sometimes it seems that the critics confuse these two senses in which one might be working on towards a concept of disease.

Paying attention to this confusion may at least partially help explaining why Boorse and his critics answer differently to the first question I raised from the critics' objection based on choice of goals. The question was:

Is there a choice of goals involved in the BST or may those goals be derived from biological theory?

The critics' claim is that since biology offers many ways of conceptualizing the facts, there is a choice involved; survival and reproduction are not the only options for goals and they are chosen over other possible ones. Boorse answers there is no choice involved; the BST's "choice of goals" is fixed by medicine's concepts of health and disease. If one wants to argue against those goals, one must argue against medicine's choice of combating "disease" as applied in medical theory.

Boorse is technically right here. If the BST is *just an analysis* of medicine's health and disease concepts, it is correct to say that the BST does not choose the criteria (statistically normal or subnormal contribution to survival and reproduction) for the *given* concepts (health and disease) it analyzes. It only makes claims about how medicine's usage of those concepts should be correctly interpreted. Consequently, to show that the BST is mistaken in its concepts or stated criteria, would be to show that medicine's usage differs from Boorse's interpretation. In this sense, it is irrelevant whether there is a choice of goals involved behind medicine's usage of its concepts. If there happens to be a choice, it is not in the level of the BST itself. Boorse's critics' answer is really on the level of medicine's usage. In the BST they have a wrong target. That is, *if* you accept Boorse's claim that he is "just analyzing" a concept construed from medical theory.

Does it matter whose choice it is?

Boorse's appeal to the claim that the BST is "just an analysis" of medical language is unsatisfactory. From the context it is clear that the reason why this conversation between Boorse and his critics is happening is not only to find out what is the correct interpretation

of medicine's usage of its concepts. The main motivation driving Boorse's work throughout the decades was not just to present medicine's health concepts as precisely as possible. The motivation was *to show that the concepts in question are value-free and scientific*. This is explicitly expressed throughout Boorse's work, notably – for example – in the first sentence of the abstract of his seminal paper (Boorse 1977). Also, in a reply to his critics, Boorse stated:

> Medicine has a distinctive theoretical foundation in a value-free science of health and disease (Boorse 1997, p.23).

In addition, many respectable writers interpret Boorse's work as aiming to show that medical theory itself– not only one concept – is at heart value-free and scientific (Fulford 2001, p.80).

Hence, the conversation is floating a bit astray if it focuses too much on the question about the level on which the choices are made. The critics are claiming that it is a value-laden choice to use survival and reproduction as criteria for defining health and disease. Boorse is saying it is not the BST's choice but concedes it might be a choice medicine makes (Boorse 1997, p.25). In the same article though, Boorse later defends the position that the stated goals – survival at least – may be derived from physiological theory[38] (Boorse 1997, p.28). If he was just analyzing medicine's use of language, why would he bother to give such a defense? And this is not anything "new". Boorse mentions his view (of functions as contributions to survival and reproduction) deriving from physiology already in the 1970s. Then he also provided a rationale for "choosing" physiology and not some other subfield of biology as the background theory for medicine. (Boorse 1977, p.556.)

All this might be relevant if Boorse wants to stick to his claim that medicine's concepts of health and disease are value-free. Even if the critics sort of missed the target here, the criticism should not be dismissed just by noting the correct target to be medicine and its supposed choices. It is then worth investigating whether medicine's concepts are value-laden in some way, which is the issue Boorse's critics are adhering and about which Boorse too seems to have opinions.

[38] Physiology is a subfield of biology.

So, Boorse has conceded that the goals might be "chosen" in a sense, but the choice is made by medicine, not by him or the BST. Secondly, he thinks medicine's choice is due to the fact that physiology seems to be the subfield of biology most closely related to individual functioning, which medicine is interested in and thus "values" in a way. This choice of physiology *is* value-laden in a sense, as medicine could have chosen something other as its subject of interest than physiology and physiological functions (i.e. "health").

Hence, the real argument between Boorse and his critics culminates in the assessment of the third question I presented earlier. The question was:

If there is a value-laden choice, what is its relevance?

The relevance of the value-ladenness in the choice-level

Let us now assume there are values involved in the choice level. Medicine has made a choice – according to its interests and values – of applying physiological theory and derived the goals of the organisms from that theory. Now, there might be many kinds of senses in which the value-ladenness of a medicine's choice of interest might be "relevant". For example, it might be socially, politically, philosophically or scientifically relevant. People care a lot about how medical practice is organized in society. If the goals medicine assumes as essential parts of its theory have some practical implications to people´s lives, then the choice of those goals (and the alleged values behind the choice) matters socially and politically. Scientifically and philosophically, it is rather obvious why it matters how the theory is constructed. Theories are fundamental subjects of study for these fields. There are probably other ways too in which the choice might be relevant, but the mentioned ones should cover most of the area of interest here.

In what follows, I try to make some sense of the conversation between Boorse and his opponents by analyzing how they seem to see this relevance-issue differently and how it affects their conclusions.

Since Boorse is primarily doing conceptual analysis, he assesses the relevance of the choice and the alleged values behind it philosophically. When answering to his critics, he strives to uncover the choice's implications for the ensuing concepts´ metaethical status, i.e., he takes as his task to determine whether the choice's alleged value-ladenness potentially implies something about the concepts' possible value-ladenness.

Boorse thinks it does not (1997, p.27 & 2014, p.692). Boorse argues that the value-laden choice of combatting disease (i.e. choosing physiology's functions and goals) would imply value-ladenness of the *concept of disease* only if the meaning of "disease" was fixed by medical practice (Boorse 1997, p.25). He means that if "disease" only meant whatever doctors were currently treating, then the whole concept would be value-laden, since treatment judgements are evaluative. It is probably reasonable to assume – as Boorse does – that treatment decisions are always at least partially evaluative. However rigorously one defines medical concepts, one must always ask: *should this condition be treated*? And the answer to this question cannot be derived from science alone.

But, as Boorse notes, it is a mistake to think "disease" as defined by medical practice. There are lots of conditions medically treated which are never described as diseases or pathological conditions by physicians: such as fertility problems or unwanted pregnancies (Boorse 1997, p.26). Thus, the concept of disease is defined separately from medical practice and the evaluative character of the latter does not make the concept itself evaluative.

Furthermore, Boorse argues there is no logical reason to think that the values involved in a choice of what to pursue would somehow taint also the concepts themselves with values. Physicians share the commitment to treating disease similarly as teachers share the commitment to spreading truth. Neither of those commitments imply that *disease* or *truth* – respectively – are normative concepts. Likewise, if you choose to build your house of wood over concrete, your evaluative choice does not make the concepts of wood and concrete value-laden. (Boorse 1997, p.27.) The fact that you evaluate wood as better than concrete, does not prevent you from giving "wood" a purely naturalistic and value-free definition. Analogically, if you think *health is good*, it does not prevent you from defining "health" in a naturalistic value-free way. As Boorse notes (2014, p.692), the definitions of many generally good things – such as wealth, social power or physical strength – are value-free. Hence, there is no reason to think that considering X worth pursuing would imply "X" is value-laden. This argument shares the logical structure with the argument stating that grounding health in physiological functionality does not make the concept of health value-laden, even if the choice of grounding health in physiological functionality is value-laden.

So, Boorse's argument is that 1) since the concept of disease is not fixed by medical practice, the presumed evaluativeness of medical practice does not imply the concept of disease to be value-laden and that 2) the fact something is thought valuable (or disvalued) does not imply it cannot be given a value-free description. Hence, "disease" can be defined value-freely in medicine, despite its status as a disvalued concept in common language. There does not seem to be anything wrong with this line of reasoning and even some of Boorse's critics concede that on this level, disease can be defined value-free (DeVito 2000, p.543; Kingma 2014, p.600). Thus, I think it is fine to conclude that Boorse succeeds in showing that *the description of the concept of disease can be given in value-free terms.*

So, the fact that values play a role in the choice-level does not make the concepts of health and disease value-laden in the level of description. Does this mean that the value-ladenness in the choice-level could not be an issue in any other way either? I think it does not. Furthermore, I do not think it is completely satisfying for Boorse to show that the description of "disease" can be given in value-free terms, despite his claims that he is only doing an analysis of medical language. Throughout his abundant work on disease, he has defended a deeper position than just the possibility of a value-free description of health and disease.

The choice of goals and the extension of disease

Recall that, in the first place, the importance of the question about the boundaries between health and disease is closely connected to the social significance of these concepts. Boorse himself provides a nice overview of the various ways in which the setting of these boundaries affects our institutional processes and their outcomes (Boorse 2011, p.15–16). Whether a condition is counted as a disease matters in many sociopolitical areas such as public and private health care, legal system and insurance.

Importantly, a major part of Boorse's original motivation in defining disease was to counter the antipsychiatrists (and normativists), whose claims were based on a presumption that 1) what we count as disease affects our social practices and 2) values play a significant role when assessing what to count as disease. To counter these claims, it is not enough to show that *after* choosing what to pursue (normal functioning relative to survival and reproduction), one can define the following concept in value-free terms. One

must be able to show that there aren't other as reasonable options available that – if chosen based on values – would lead to different sets of conditions labeled as diseases. If there is such an option available and the choice between the options involves values, then I do not see why the possibility of a value-free description of the concept resulting from that choice would free us from considering the effects of our values behind the choice.

The critical issue is whether the choice of physiology and its implied goals (survival and reproduction) is "a value-laden move", as DeVito (2000, p.544) suggests, *and gives us a different extension of disease than some other possible choice*. If it does, it has straightforward implications for our institutional practices. For example, health insurance may exclude conditions not considered diseases (Boorse 2011, p.15). Criminals may be excused in the court of law if they are considered mentally ill (Boorse 2011, p.16). So, what becomes excluded and who becomes excused depends at least partly on the definition of disease. The theoretical definition of disease plausibly has some effect on what conditions become considered illnesses in practical contexts. Thus, the original choice of goals is by no means automatically "trivial", as Boorse seems to imply by emphasizing the point about the possibility of a value-free description of the disease concept in his defense. It may be trivial only if the choice made is the *only reasonable choice, given* medicine's mission of addressing people's physical (and mental) problems. If there are other reasonable options available and the choice affects which conditions Boorse ends up describing in value free terms, the choice matters.

Sometimes Boorse seems to *presume* there exists unambiguous and uncontroversial naturalistic extensions of health and disease and thus the only "value-laden move" that is made would be medicine's decision of promoting the former and curing the latter. It is like Boorse is saying that it was known from the beginning exactly what "health" and "disease" pick out in the nature and the choice was only about what to do with them. But this cannot be the case since the serious work towards delineating what exactly falls under those concepts only started in the 1960s (Scadding) and Boorse himself became one of the main contributors to this project in the 1970s. Much of Boorse's work through 1975–2014 is precisely about arguing what kind of conditions should or should not count as diseases. Nothing seems to imply that the notions of health and disease and their extensions have been uncontroversial and agreed upon. They certainly were not in the 1960s when this conversation started and still aren't, as the on-going quarrel shows. Medicine

may have had a customary way of using the concepts, but it does not mean that the true meaning of the concepts was uncontroversial.

Boorse cannot dodge the problem by pointing out that the concepts already exist in medical texts or that they can be given value-free descriptions. The question his critics are asking is not about *de facto* existence of those concepts in medical language or about the possibility of giving them value-free descriptions. It is about the role of values in constructing, i.e. "choosing" those concepts.

Next, I will look first into what other options have been suggested in the place of physiology and its implied goals of organisms, survival and reproduction. Then I will assess Boorse's justification for the choice of survival and reproduction as goals. Finally, I will examine whether this justification is value-free, and if not, what sort of value-ladenness it might imply in the resulting concepts.

Alternative goals

Given everything said in the last section, it is still unclear what the alternative reasonable ways of conceptualizing health and disease and "choosing" the goals of organisms' would be, at least if one wants to define concepts of health and disease so that they capture the meaning of the current concepts used by either medicine or by laypeople in any relevant ways.

Even if one is a strict revisionist like Boorse, one probably wants to keep at least *some* connection between the meanings of the established concepts and the meanings of the revised concepts if possible. If one would not see anything common between them and both were still referred to with the same term, it would seem like there is no reason not to label one of the concepts with another term and completely cut off the remaining superficial and misleading connection[39]. Thus, when trying to define somewhat established concepts, there are limits within which one should operate.

What then, would be the plausible alternatives for goals, other than survival or reproduction? Firstly, the interesting potential alternatives in this context are those which are

[39] As I noted earlier, Boorse in fact *did* so by replacing *disease* with *pathological condition* (Boorse 2011, p.26). But this probably does not solve much since Boorse still means roughly the same with *pathological condition* as he did with *disease* and most writers still use *disease*. What is considered *pathological* may have much of the same relevance as I argued what is considered *disease* as having, even if it makes the linguistic expression clearer and better in line with medical writing.

grounded on some scientific theory. The ones explicitly evaluative, such as *social well-being* will not do. If one wants to argue that "health" and "disease" *should* be overtly evaluative concepts grounded on evaluations of how well an individual's life is going on socially, it is totally fine. But here, I am trying to figure out whether the BST is value-free (or only "trivially" value-laden) as Boorse claims, not so much whether a normative account of health would be better than a value-free account. Hence, the interesting question here is whether there are plausible alternative *naturalistic conceptualizations* of health and disease. If there are such alternatives, one could perhaps claim that the choice between them is a "value-laden move" as Devito (2000, p.544) does.

Unfortunately, there aren't any thoroughly analyzed alternative naturalistic candidates around[40], as the critics have concentrated more on finding the flaws in Boorse's account than providing well-argued alternatives. Mostly, critics just mention some potential alternative goals which in principle could be "chosen", but do not provide an argument for choosing them. The critics seem to assume that it is problematic enough for Boorse's account's value-freeness to show there exists potential alternatives. For example, DeVito (2000, p.542) says that medicine could "rely upon the interests of patients, philosophers or evolutionary biologists" (whatever those interests may be) instead of the interests of physiologists. DeVito does not give any argument for choosing any of those he mentions. Moreover, at least some writers have speculated *species survival* and *ecological equilibrium* as alternative goals (Boorse 1997, p.25).

Engelhardt has suggested that to keep the definition of disease value-free, one could choose *species survival* as a goal. Then "all judgements about disease and health would simply presuppose this and no further value-judgements". (Engelhardt 1976, p.265.) The problem of species survival as a goal is that it would contradict the goal of individual survival, since there are conditions like sickle cell disease, which may be harmful to individual survival in certain environments but increase species adaptability (Engelhardt 1976, p.266). Since medicine has an explicit interest in individuals, the BST cannot choose species survival over individual survival if it aspires to analyze individual health and disease and the medical usage of those concepts.

[40] The only competent one, Wakefield's Harmful Dysfunction Analysis, applies survival and reproduction as goals. (Wakefield 1992).

In fact, Engelhardt mistakenly thought that Boorse *does* apply species survival as a goal in the BST and Engelhardt was arguing against that choice, appealing to medicine´s interest in individuals (Boorse 1997, p.24). Funnily, while intending to argue against Boorse and the BST, Engelhardt accidentally provided a defense for the BST's choice of individual survival as a goal by giving a reason why the BST should choose individual survival over species survival, as it in fact does.

The point to be taken here is that Engelhardt seems to agree with Boorse that *it is justifiable* for the BST to choose individual as the subject of interest rather than species or some other class or system (e.g. ecology). For the BST it is justifiable, because the BST is analyzing medical language and from medical texts it is clear that the interest is in individuals. For medicine, Engelhardt thinks individual is the correct unit of interest, as biological functions work to preserve the individual and not species (Engelhardt 1976, p.266).

So, while aiming to criticize the BST, Engelhardt actually managed to provide some defense for it. Furthermore, the logic just applied above should apply to all potential suggestions like *species survival* or *ecological equilibrium*, which would involve something other than individual as the subject of interest. If it is accepted that individual is the only reasonable unit of interest for medicine, then it is clear why medicine should choose physiology as its background theory rather than any other subfield of biology focusing on other biological systems than individuals. Surely, this does not counter any arguments suggesting other kinds of individual goals than those of biological nature. The argument here is only that should one accept individual as the subject of interest, then physiology is a reasonable choice over those subfields of biology focusing on other systems than individuals.

As the literature seems quite shallow in its offering on the analysis of naturalistic candidate conceptualizations, I suppose it is now better to move forward and have a look at Boorse's justification for choosing survival and reproduction as goals.

The case for survival and reproduction

Boorse noted already in 1976 that the justifiability of any goals depends on the goal-directed system one is studying. If one is studying human physiology, it is logical to take the system's of interest (individual organism) goals to be those of physiology, i.e. survival

and reproduction. If one is studying species, it is logical to take the system's (species) goal to be species survival. If one is studying ecology, it is logical to take the system´s (ecosystem) goal to be ecological equilibrium. (Boorse 1976, p.84.) Different subfields of biology have different interests and may thus use different goals as the focus of their function statements. But as *it is only the subfield of physiology whose functions seem relevant to (individual) health*, it is appropriate for medicine to ground itself in physiology and adopt physiology's goals. (Boorse 1977, p.556.)

In a very trivial sense this is still a "value-laden choice". To choose individual physiological functioning and individual goals as the main interest of medicine over, e.g., evolutionary biology and species goals or ecology and ecological goals, reflects interests and values in a very basic level. But how problematic is this kind of value-ladenness? It certainly is not equivalently problematic as labeling soviet dissidents or fleeing slaves as *mentally ill*, is it (Wakefield 1992, p.373)?

For the matter of species survival, Boorse speculates that perhaps aging may be a process which sacrifices individual survival to species survival. And if it does, medicine might be forced to make a value-laden choice between the goals (individual or species survival). (Boorse 1997, p.28.) The point is that if aging is a process sacrificing individual survival to species survival, then there is evidence of a biological function contributing to species over individual and thus a case for alternative biologically grounded concepts of health and disease. In this case, one could claim that there could be a reasonable disagreement about values between those valuing individual survival and those valuing species survival. The choice between these naturalistically grounded goals would not then be "trivial". But for the time being, since there still isn't much credible evidence for organs or processes contributing primarily to species survival at the expense of individual survival (Boorse 1997, p.28) and there is lots of evidence of organs and processes contributing to individual survival, it is not unreasonable to suggest that the choice between the individual or species goals remains "trivial".

Survival seems to be an undisputed goal of individual organisms both in a naturalist sense (Boorse 1977, p.556) and in a common sense – almost everybody wants to stay alive (Boorse 1975, p.60). Nobody is toying with the idea that perhaps we should consider *death* instead as a goal and educate physicians to promote death. Nobody seems to even suggest that we should be *indifferent* about life and death. The *value of life*, and thus *survival*, seems to be shared by all humans.

Boorse concedes reproduction is more controversial because it sometimes conflicts with individual survival – i.e. pregnancy and childbirth increase mortality (Boorse 1997, p.28), but it is nonetheless a capacity which seems an essential part of our species' physiological design (Boorse 1977, p.556). And since no equally clear examples of organs or processes sacrificing the individual to species (or ecological) goals have been found, medical disease concept might (at worst) be indeterminate as to goal (Boorse 1997, p.28).

Reproduction may be seen problematic as a goal also because it is so often deliberately prevented by using contraception. It seems that lots of people are striving to enjoy themselves sexually, while taking elaborate precautions to *avoid* reproduction (Kingma 2014, p.601). Thus, one might ask, why should we consider *reproduction* but not *enjoying sex* as a goal of an organism? If one desires sex and childless life, aren't these desires as "biological" as an urge to reproduce? This is a good question. As Boorse has conceded that reproduction is more problematic as a goal than survival, it is unsettled whether choosing reproduction as a goal could be claimed "trivial".

While I think Kingma's point about sex and reproduction is worth attention, I consider it a bit misleading to frame the use of contraception as evidence of a "desire to avoid reproduction" as she seems to do (Kingma 2014, p.601). Contraception and other means of birth control are mostly about *delaying* reproduction, not about abstaining from it altogether (most people still reproduce). People have limited ability of generating offspring and limited resources to care for them. Hence, *given* a desire to reproduce, it is still logical to use contraceptive measures as means of timing one's reproduction.

It is true that some people choose not to reproduce at all. But it is possible that they still appreciate *reproductive capacity* and it is just the various contingent factors that finally tilt their choice towards staying childless. Note here that it is really the *capacity* to function within normal variation in which the BST's notion of health is grounded (Boorse 2014, p.715). Physiologically, what matters for disease judgements, is whether an individual *could* reproduce at a statistically normal efficiency. Whether one *actually* produces offspring or not does not matter for disease judgements.

Perhaps there are also people who won't appreciate even the capacity to reproduce and wish to get rid of it. It is plausible there exist some such people. But if they are rare

enough, I do not think their existence would hurt the BST much. The "failure"[41] to appreciate reproductive capacity could be interpreted as a deviation from normal human functioning and as such a "disease" in the BST's technical sense. Perhaps, it would be labeled as a *mental disorder*, as it is about "subnormal" desires. Then, it *would* depend on values whether medicine would have any interest in getting involved in the treatment of that kind of deviation from normal functioning. But that would not bother Boorse as he is not interested in *medical practice*.

Hence, I do not see reproduction as problematic as a goal as Kingma sees it. Kingma's (2014, p.601) argument against reproduction seems partly based on the observation that trying to avoid reproducing is a "common desire" among humans. As I argued, this observation confuses timing or delaying reproduction with avoiding it. The remaining problem is that even if reproduction would be assumed as a plausible goal, there might be other as plausible options and the choice between them could (in principle) be value-laden, as the critics claim.

The central point of Boorse's argument

What Boorse is saying is that the appropriateness of a goal depends on the context of interest. As medicine's context is addressing human physiological (and perhaps mental) problems, it is reasonable to apply physiological goals in medical theory. It is not sensible to claim that *given* the objective of addressing physiological problems, one could as plausibly choose "species survival" or "ecological equilibrium" or "social prestige" as a goal for the contributions of organism's part-functions. The choice that is made by medicine, according to Boorse, is to address the individual physiological problems people are demanding to be addressed. There is a value involved in choosing to answer to people's needs, but that value does not make the concepts of health and disease value-laden in a problematic way.

I conclude that because writers criticizing medicine and the BST of the value-laden choice of goals have not presented a plausible naturalistic alternative for physiology and its implied goals, it is possible that Boorse is onto something when claiming that medicine's and thus also the BST´s choice is trivial. In addition to the evident lack of alternatives, Boorse has given quite a carefully constructed rationale for medicine's choice which he

[41] I.e. "not appreciating" in descriptive language.

applies in the BST (Boorse 1976 & 1977). It is possible that the value Boorse admits being behind the choice (value of physiological functionality) is shared enough as to make the choice non-problematic as Boorse claims. Why wouldn't medicine strive towards promoting "health" as physiological functionality if that is what people unanimously want and there does not seem to be plausible alternatives on offer?

The choice of promoting physiological functionality, as well as the concept of disease, would still be value-laden in a trivial sense: physiological functionality (which itself can be defined value-free) is studied and promoted by medicine, because everybody likes physiological functionality. Disease and its extension may be different with the current choice than they would have been, had medicine interested itself with something other than physiological functionality. "Disease" would then pick some other set of conditions than it currently picks. But if the preference for physiological functioning is *universally shared*, it does not seem to be very condemning to say that choosing (scientific field's interests) according to that universal preference is value-laden. The shared values related to physiological functioning imply that the extension of disease based on physiological functions should[42] be roughly what people want. And the fact that what is accepted as disease in somatic medicine is mostly uncontroversial hints that those values truly are shared. In contrast, the stark social, political and academic controversies around the concept of *mental disorder* might suggest the opposite; the values related to mental functioning (our beliefs, emotions, desires etc.) might be *diverse* (Fulford 2001, p.82). Everybody thinks cancer and broken bones are bad but only some percentage of people think sexual fetishes, hyperactivity or even delusions are unquestionably bad.

The question of value-ladenness of scientific concepts is difficult to deal with largely because there is no shared understanding of the role values play in scientific processes and theories[43]. Still, I think Boorse has made a decent case here and it is the critics' burden to provide an argument for reasonable alternative goals, which existence would make medicine's "choice" more dubious.

[42] In the sense of "probably is".
[43] See Kincaid et al. (2007) for an analysis of roles of values in science.

6.2 The choice of reference classes

In the Biostatistical Model, the term "statistical" implies that the boundaries between normality and abnormality – i.e. between health and disease – are determined statistically. This means that it is assumed that a certain range of variation in a trait is "normal" and if the trait functions below (or above) this normal range, it is (usually) considered subnormal and thus pathological. The assessment of the variation falling into the normal range must be done in relation to some population of individuals.

In the BST, an organ's normal function is a statistically typical contribution by it to individual survival or reproduction of the member of the *reference class* (Boorse 2014, p.684). A reference class picks out the relevant population in relation to which the statistical normality of function is assessed. In the BST, the reference class Boorse uses is *an age group of a sex of a species*[44]. According to Boorse, this "choice" is connected to species design, is derived from biological theory and is consistent with medical usage (1977, p.557–558; 2014, p.695).

Hence, what is considered normal within the BST's framework, depends on the reference class in question. For example, the normal range of heart functioning depends on whether the subject examined is a young woman or an old man. The "efficiency" which is normal for an old man's heart may be considered subnormal and thus pathological for a young woman.

This creates a logically similar problem as the one related to the choice of goals discussed in the last section. The reference classes Boorse applies may be seen as a result of a "choice" and what is chosen may affect who becomes considered diseased (Kingma 2013, p.369–370). Then it may be asked whether there is some value involved in the choice process that led Boorse to adopt the current reference classes. If there is, it may be claimed that the ensuing concept of disease is value-laden.

Kingma (2013, p.369–370) argues that if different reference classes were adopted, then different accounts of function, health and disease would emerge. Kingma's example illustrates this neatly:

> Suppose the XST is an account of disorder that closely resembles the BST, with one exception: sexual orientation is a reference class on the XST.

[44] Boorse also considers the possibility of factoring in "race", since different races have different functional designs (1977, p.558).

> According to the BST homosexuality is a disorder, because non-heterosexual attraction is a worse-than-typical contribution to reproduction. On the XST, however, homosexuality is normal in the reference class of people with homosexual orientation, and therefore healthy. (Kingma 2013, p.370.)

Kingma's example shows how making one more subdivision to the system of reference classes changes what is considered normal function, health and disease. While in the BST homosexuals' reproductive functioning would be considered subnormal and thus pathological, in the XST the same functioning would be considered normal since it was assessed relative to a reference class consisting only of homosexual individuals. In relation to a group of homosexuals, individual homosexual's reproductive functioning is normal.

Once again – as in the case of the choice of goals – the extension of "disease" seems to depend on how the choice is made. In this case, it depends on what are chosen as reference classes. Had the reference classes been chosen differently, some other set of conditions would have been classified as diseases. And if this supposed choice was based on underlying values, then the ensuing concepts of health and disease could be claimed to be value-laden.

Kingma argues that Boorse needs to provide a non-circular justification for his proposed reference classes: why age and sex but not some alternative way of grouping people? This justification should be presented in a value-free manner, as Boorse is committed to arguing that the preference for the BST over other options represents a value-free choice. But Kingma does not think such a non-circular, value-free justification is possible. It would be circular to restrict reference classes to healthy groups of humans. (Kingma 2013, p.370.) If one argues that "homosexuals" cannot be a reference class, it would seem that one has already decided what is "normal", i.e. heterosexuality.

In the absence of any objective justification for choosing the BST over the XST, there is no way of settling the demarcation problem regarding homosexuality (Kingma 2013, p.370). To put it more clearly, Kingma means that one cannot say whether homosexuality is a disease or not, if one has no objective grounds for choosing the system of reference classes. And since XST-type accounts can be generated for many if not all "disorders" (Kingma 2013, p.370), the problem can be generalized: there is no way of settling the

demarcation problem for almost any conditions, since one can always suggest a new reference class in which the condition in question represents normal functioning.

In even more general terms, *a naturalist* must demonstrate a value-free justification for employing the particular naturalistic concept he has chosen rather than another one (Kingma 2013, p.370). That is, if the natural world can be divided in conceptual pieces in various optional ways, a naturalist must offer some rationale for choosing one particular way. Even if it is conceded that it is okay for a naturalist to say that he is interested in disease because disease is disabling or painful, he must provide a value-free justification for favoring one possible naturalist extension of disease over another (Kingma 2013, p.371).

The criticisms concerning the choice of goals and the choice of reference classes have the same logic at their roots. If there are optional ways of choosing goals or reference classes and the choice affects the extension of disease, one should provide a value-free justification for the preferred choice. That is, if one wants to keep the ensuing concepts of health and disease value-free. After going through the analysis of the criticism concerning the choice of goals, it is not difficult to guess how Boorse might answer to this criticism concerning the choice of reference classes.

Boorse's counter-argument goes as follows: the BST analyzes "health" using the concepts of *statistical normality, survival, reproduction, organism, part, process, species, sex, age* and *causation*. None of these concepts are alleged to be value-laden even by the BST's critics. (Boorse 2014, p.693.) Hence, as health and disease are defined in value-free terms, the concepts themselves are value-free.

Furthermore, as in the case of the goals, Boorse claims he does not "choose" the reference classes, the choice is medicine's. He tries to choose the analysis which best fits biological theory and medical usage (Boorse 1977, p.551; 2014, p.693). Boorse notes that his preferred reference classes are almost straightforwardly derived from biology:

> Apart from one detail, the BST's reference class is just one morphological type in the smallest taxon to which an organism belongs. When biologists describe and classify organisms, they sort them into species or subspecies, separate them by sex in sexual species, and distinguish immature from adult forms. The BST departs from this only in making old age as well as youth count for classification… (Boorse 2014, p.695.)

And again, as in the case of goals, we arrive to the question about the rationale for the supposed "medicine's choice". *Given* people's shared interest in their physiological functioning, is it justifiable – or "trivial" – to use age and sex (derived from biological theory[45]) as reference classes or are there other reasonable options that could have been chosen as well? Kingma argues that we could easily construct a new reference class for groups of people with any kind of conditions and thus the BST would seem arbitrary with its choice of sex and age. In contrary, Boorse thinks it is sufficient that the BST, with sex and age as its reference classes, agrees with biological theory and medical usage, the latter being the one the BST means to analyze. If one can derive the preferred reference classes from biological (physiological) theory and there are no *reasonable alternatives,* then one may conclude that medicine's "choice" of those reference classes is the only logical one and not value-laden in any meaningful way.

Th core questions here are the same I have raised in the last section concerning the choice of goals. If the choice of reference classes affects the *extension of disease* and if that extension matters socially, politically etc., isn't it important what becomes chosen and how it is chosen, regardless of whether the choice is made by Boorse or medicine? If the choice is based on values, aren't the ensuing health concepts value-laden in a way that could be worth attention? And shouldn't Boorse also be concerned with these questions, as in many occasions he seems to defend the value-free scientific status of medicine's health concepts in addition to "only analyzing" them?

I addressed similar questions in depth in the last section and commented them also in this section, so I will not go through them again. I conclude that, first, Boorse *should* address these questions (as he has done) and second, the answers depend roughly on whether the choice of reference classes may be reasonably considered as "trivial", *given* people's shared interest in physiological functioning and *given* biological theory provides us an uncontroversial way of determining the appropriate reference classes. In the case of choosing goals, I also concluded that Boorse in fact has given a well-argued rationale for thinking that the choice truly is trivial. Here, I make the same conclusion. Boorse has argued for the choice of reference classes already in his old theory paper (1997, p.556–558) and defended this position recently (2014, p.692–694) in a credible way. If the critics

[45] For Boorse's detailed explanation on why specifically sex and age generate such distinct species designs that they should be used as reference classes, see Boorse (1977, p.558).

wish to object, it is not enough to point out the existence of alternative options for reference classes. They must show that there are alternatives which are as plausible as the BST's reference classes, given biological theory and popular interest in physiological functioning.

7 Conclusions

7.1 Challenges of interpretation

Christopher Boorse has written about health and disease for four decades, beginning from his first seminal paper (Boorse 1975) and finishing with the Second Rebuttal (Boorse 2014) for the time being. During the decades, he has written at least 13 papers related to health concepts (Boorse's CV, 2019), some of them quite comprehensive. While 13 papers in four decades may be considered infrequent, it is still a respectable amount of work.

The sheer amount of publications and the fact they are spread over four decades may be causing some of the confusion in the debate. Occasionally, a critic seems to have read certain selected papers and based his/her interpretations of Boorse's arguments on those papers. Then years or perhaps a decade later, Boorse might reply that the critic's arguments lead astray, because they are based on misdescription or misinterpretation of the BST[46].

It is difficult to say whether the issue on these occasions was the negligence of background work by a critic or ambiguousness and/or incoherence between Boorse's papers from different periods. My reading is that depending on the occasion, both have played a part.

Four decades is a long time and understandably some of Boorse's views have changed[47] on the way. Also, his emphasis has changed from mental disorders to somatic disease (Boorse 2014, p.691). Lastly, as I have argued, it is evident that Boorse has not *only* worked on the correct interpretation of medical language, as he sometimes claims. He has

[46] For example, in his Second Rebuttal (2014), Boorse lists Bolton, Murphy, Ananth, Venkatapuram, Budson, Richman and Cooper as writers misdescribing or misinterpreting his arguments and the BST. Then he refers his own work from the 1970s till 2014 to prove his case. (Boorse 2014, p.686–691.)

[47] For example, his views on correct terminology and on the usefulness of the distinction between disease and illness (Boorse 2014, p.684).

also defended the notions of health and disease as fundamentally *scientific and value-free concepts*.

The changes of views, the change of emphasis and the ambiguousness of Boorse's aims all contribute to the challenge of interpreting his work correctly. Thus, it is not surprising there is some confusion in the extensive debate I have presented in this thesis. In the end, I hope I have managed to make sense of at least some of it and succeeded in steering the focus on the important issues. But as more qualified writers have been led astray, it is probably good to approach the interpretations I have made with some caution too.

7.2 Limitations

The difficulties with the interpretation of the literature point to the possibility that some of my analysis and conclusions may be incorrect or past the subject. Whether that in fact is the case, may only be settled in future discussion. There are other things though, which more clearly limit the implications of my analysis.

The obvious shortcoming of my research project is that it is limited exclusively to the Anglo-American context. Practically all the relevant works I have utilized during this project are from native English-speakers. In the literature about health concepts I have reviewed, there are very rarely mentioned any works from people outside of the Anglo-American sphere. This is certainly understandable, since it is easier to operate within a linguistic community than between two. Nonetheless, this surely is a limiting factor. Within other linguistic communities, there might exist significantly different conceptions of health and disease about which I am totally unaware of. Moreover, as far as any implications for institutional practices are concerned, they might change a lot in different cultural and institutional contexts.

Another factor limiting the generalizability of inferences drawn from my work is that the study was centered on a particular conversation which started from the controversies related to psychiatry and has then revolved mostly around Christopher Boorse's Biostatistical Model. The second most influential (partly) naturalistic account of disease is Jerome Wakefield's Harmful Dysfunctional Analysis and except few remarks, it was ignored in this thesis. This was a choice I made as I judged that Wakefield's model is similar enough to Boorse's in its naturalistic details, so that including it would mostly just make the analysis heavier. Still, there are some important points that could be addressed by comparing

Boorse's and Wakefield's models[48]. For example, the relationship between theoretical and lay concepts is one. The question about the consequences of including *harm* in the definition of disease is another one. Moreover, there is a parallel discussion of health concepts going on in the context of *disability studies* (see Vehmas & Mäkelä 2008a; 2008b). Other similar parallel discussions may exist which could be relevant for the topic. Exploring these potential discussions is left for future work.

As I have emphasized from the beginning, my analysis concerns primarily the theoretical concept of disease and presumes health is just absence of disease. Hence, most I have said about the value-ladenness of disease applies only to this narrow kind of theoretical concept. It is a separate – though related – conversation, whether and on which level(s) "disease" (or illness or disorder) as used in practical contexts is value-laden and what it may imply. My analysis also ignores any positive conceptions of health.

7.3 Results

Keeping in mind the limits related to the scope of my analysis and the uncertainties due to difficulties in interpretation, I will now sum up the observations I have made and which I think may be of some value in the future.

In the beginning of this essay, I pledged to search an answer for two interconnected questions. First, the aim was to assess whether disease is a value-free or a value-laden concept. Then, should the concept turn out value-laden in a certain suggested way, the purpose was to elaborate what it means. I believe there is something I can contribute to both questions.

Let's start with the points I feel quite confident making. The easiest one is the following: should one accept the BST as *just a description* of the way "disease" is used in medical theory, then the question about the BST's value-ladenness would appear pointless. If one aims just to describe how other people use language, he might hit or miss, his motives might be value-laden, but there is no reason to think the resulting description itself would be value-laden in any interesting sense. As I have argued, I do not accept Boorse's claim that he is just describing or "analyzing" medical language. I believe it is a reasonable interpretation of his work that he is also *defending medicine as a scientific and value-free*

[48] For the comparison, see Kingma (2013) and Wakefield (2014).

discipline (Boorse 1997, p.23). So here I am with the critics and I do not think framing the BST as "just an analysis" is enough for evading the questions concerning value-ladenness.

The second point is somewhat related to the first one. Fulford (2001) argued that in constructing the BST and defining "disease", Boorse uses value-laden terms, which on Fulford's interpretation shows that disease and the BST are value-laden. Boorse (1997) has explained why it is possible to replace all the seemingly evaluative terms in his model with clearly descriptive terms and thus show that the model itself and "disease" are both value-free. As I argued, Fulford mistakenly presumed Boorse aiming to give an evaluative concept of disease a value-free description. This is not what Boorse is doing. He aims to show that the concept *is* value-free by giving it a value-free description. Fulford misinterpreted Boorse and misses the point. I think Boorse is right that after choosing what to analyze or pursue – irrespective of whether the choice itself is value-laden or not – the target concept can be given a value-free description[49]. This is the case with concepts such as "wealth" or physical "strength" so why not also with "health" and "disease"? Thus, the BST is not value-laden in the level of description.

The most complex challenge I had to tackle in this project was to address the possible value-ladenness of the BST in the choice-level. Given different values, could the reference classes or the goals of the organisms have been "chosen" otherwise, and if yes, would that imply that the BST and the concept of disease it describes are then value-laden? Without repeating the complicated discussion on this topic, I conclude that yes, the goals and the reference classes could have been chosen otherwise and in this sense, it is not necessarily wrong to call the BST value-laden.

However, just noting that the BST might be value-laden in the choice-level does not tell much. The interesting task lies in elaborating the meaning of this kind of value-ladenness. As I argued, Boorse has made a credible case for the current goals and reference classes of the BST. According to the BST, medicine has derived the reference classes and the goals from the appropriate biological theories, *given* the common interest in human physiological functioning. Should someone argue for scientifically better choices given this common interest, medicine – and the BST – might choose again. Then we might have a

[49] Many critics of Boorse also agree on this point, e.g. DeVito (2000, p.542).

bit different concept of disease than we currently have. This seems like a standard scientific practice.

There are always some fundamental values driving human enterprises and the values may affect what is pursued. On this level it may be that those enterprises – such as medicine – are value-laden. But if the fundamental values driving medicine and its choices are shared enough, it may be the case that medical theory, the BST that aims to track it and the concept of disease are value-laden only in a very trivial and inconsequential sense. Perhaps value-ladenness becomes a crucial issue only in contexts where values are diverse and competing against each other. In those kinds of contexts, it could matter a lot if an enterprise is laden with particular values instead of other values.

Values are involved in the science of medicine, but it does not imply there is no science or that the science is bad. Moreover, the fact that values play a role in the choice-level does not *necessarily* imply any normative flaws in medicine, in the BST or in the concept of disease. Whether such flaws are implied, requires further research.

7.4 Suggestions

In the light of the observations I have made, I think it is reasonable to conclude that research into the questions about value-ladenness of health concepts should further shift focus on the role and meaning of values. It is all but settled that on some level(s), values have a role in science and in the construction of scientific concepts, such as disease. At the same time, it has become evident how weak and inconsequential it is just to note that a scientific discipline or a scientific concept is value-laden. The meaningful questions concern the level on which the values enter and the role they play.

Promising work on these kinds of questions has been recently done in the context of health concepts by Elselijn Kingma who has distinguished different domains where naturalist and normativist claims can be contrasted (Kingma 2014). She has also suggested that the way forward is to combine insights from both camps instead of polarizing the dichotomy further (Kingma 2017). I eagerly agree. Related ideas have been proposed by Broadbent (2017). Broadbent suggests that instead of representing opposite poles of unidimensional debate, naturalism and normativism are occupying two diagonally opposing positions in a two-by-two matrix, where dimensions are objective vs. subjective and value-free vs. value-laden. This kind of fine-grained analysis brings nuance into the debate. Perhaps

combining the approaches proposed by Kingma and Broadbent with the general analysis of the role and meaning of values in science by Kincaid et al. (2007) would provide effective tools for progress.

References

Agich, G. (1983). *Disease and value: A rejection of the value-neutrality thesis*. Theoretical Medicine, 4(1), p. 27–41.

American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.).

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th ed.)*.

Bickenbach, J. (2017). *WHO's Definition of Health: Philosophical Analysis*.T. Schramme, S. Edwards (eds.), Handbook of the Philosophy of Medicine, Springer Science+Business Media Dordrecht 2017. 961–973.

Boorse, C. (1975). *On the Distinction Between Disease and Illness*. Philosophy and Public Affairs 5:49–68.

Boorse, C. (1976a). *What a Theory of Mental Health Should Be*. Journal for the Theory of Social Behaviour 6, 61–84.

Boorse, C. (1976b). *Wright on Functions*. The Philosophical Review, Vol. 85, No. 1 (Jan., 1976), p. 70–86.

Boorse, C. (1977). *Health as a Theoretical Concept*. Philosophy of Science, Vol. 44, No. 4 (Dec., 1977), The University of Chicago Press on behalf of the Philosophy of Science Association, p. 542–573.

Boorse, C. (1997). *A Rebuttal on Health*. In What Is Disease?, eds. J. M. Humber and R. F. Almeder, 1–134. Totowa, NJ: Humana Press.

Boorse, C. (2014). *A Second Rebuttal on Health*. Journal of Medicine and Philosophy 39:683–724.

Boorse, C. (2019). *Curriculum Vitae*. Retrieved 30.1.2019. https://www.philosophy.udel.edu/Documents%20Bios%20CVs/Chris_Boorse_CV.pdf.

Broadbent, A. (2017). *Health as a Secondary Property*. Brit. J. Phil. Sci. 0 (2017), 1–19.

Brown, W. M. (1985) *On Defining ´Disease´*. J. Med. Philos. 10, 311–328.

Callahan D (1973). *The WHO definition of 'health'*. Stud Hastings Cent 1:77–87.

Canguilhem, G. (1991). *The normal and the pathological* (1st paperback ed.). New York: Zone Books.

Cooper, R. (2002). *Disease*. Stud Hist Phil Biol Biomed Sci 33:263–282.

Cooper, R. (2015). *Must Disorders Cause Harm? The Changing Stance of the DSM*. History, Philosophy and Theory of the Life Sciences, Vol. 10. The DSM-5 in Perspective: Philosophical Reflections on the Psychiatric Babel. 83–96. Ed. Steeves Demazeux & Patrick Singy.

Decker, H. S. (2013). *The Making of DSM-III: A Diagnostic Manual´s Conquest of American Psychiatry*. Oxford University Press. New York.

DeVito, S. (2000). *On the Value-Neutrality of the Concepts of Health and Disease: Unto the Breach Again.* Journal of Medicine and Philosophy 2000, Vol. 25, No. 5, pp. 539–567.

Engelhardt, H. T. (1976) *Ideology and Etiology*. J. Med. Philos. 1, 256–268.

Ereshefsky, M. (2009). *Defining 'health' and 'disease'*. Studies in History and Philosophy of Biological and Biomedical Sciences 40: 221–227.

Eysenck, H. J. (1960). *Classification and the Problem of Diagnosis*. In Handbook of Abnormal Psychology (ed. Eysenck) London.

Foucault, M. (1965). *Madness and Civilization: A History of Insanity in the Age of Reason*. (R. Howard, trans.). New York: Pantheon. (Original version published 1964).

Fulford, K. W. M. (1989). *Moral Theory and Medical Practice*. Cambridge, New York.

Fulford, K. W. M. (2001). *´What is (mental) disease?´: an open letter to Christopher Boorse*. Journal of Medical Ethics 2001; 27:80–85.

Garson, J. (2016). *A Critical Overview of Biological Functions*.

Gavin J.M., Kwoh, N.Y-L.; Jo E., Liang M.T.C. (2015). *Low Body Mass Index and Osteoporotic Fracture Risk in Young Women*. Medicine & Science in Sports & Exercise. 47(5S) (Supplement 1): 617–618, May 2015.

Goffman, E. M. (1961) *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates.* New York. Doubleday.

Goffman, E. M. (1963). *Stigma: Notes on the Management of Spoiled Identity*. Englewood Cliffs, NY: Prentice-Hall.

Guerrero, J. D. (2010). *On a Naturalist Theory of Health: a Critique.* Studies in the History and Philosophy of Biological and Biomedical Sciences 41, 272–278.

Guze, S. B. (1978). *Nature of Psychiatric Illness: Why Psychiatry is a Branch of Medicine*. Official Journal of the American Psychopathological Association, Vol. 19, no. 4, 295–307.

Haslanger, S.A. & Saul, J. (2006). *Philosophical analysis and social kinds. What good are our intuitions?* Proceedings of the Aristotelean Society. Volume 106, Issue 1, January 2006, Pages 89–118.

Hinderliter, A. (2015). *Sexual Dysfunctions and Asexuality in DSM-5*. History, Philosophy and Theory of the Life Sciences, Vol. 10. The DSM-5 in Perspective: Philosophical Reflections on the Psychiatric Babel. 125–139. Ed. Steeves Demazeux & Patrick Singy.

Iyengar N.M., Brown K.A., Zhou X.K., et al. (2017). *Metabolic Obesity, Adipose Inflammation and Elevated Breast Aromatase in Women with Normal Body Mass Index.* Cancer Prev Res (Phila). 2017;10(4):235–243.

Kataoka, H. and Ueno, S. (2017), *A giant uterine tumor in a woman with myotonic dystrophy*. Clin Case Rep, 5: 539–540.

Kendell, R. E. (1975). *The Concept of Disease and Its Implications for Psychiatry*. British Journal of Psychiatry, 127, 305–315.

Kendell, R. E. (2001). *The Distinction Between Mental and Physical Illness*. British Journal of Psychiatry, 178, 490–493.

Kincaid, H., Dupré, J. & Wylie, A.(eds.) (2007). *Value-Free Science?: Ideals and Illusions*. Oxford University Press,. Retrieved 25 Jan. 2019, from http://www.oxfordscholarship.com.libproxy.helsinki.fi/view/10.1093/acprof:oso/9780195308969.001.0001/acprof-9780195308969.

Kingma, E. (2010). *Paracetamol, Poison, and Polio: Why Boorse's Account of Function Fails to Distinguish Health and Disease*. British Journal of Philosophy of Science 61 (2010), 241–264.

Kingma, E. (2013). *Naturalist Accounts of Mental Disorder*. The Oxford Handbook of Philosophy and Psychiatry. Chapter 25. 363-384. Ed. K.W.M. Fulford, Martin Davies,

Kingma, E. (2016). *Situation-Specific Disease and Dispositional Function*. Brit. J. Phil. Sci. 67 (2016), 391–404.

Kingma, E. (2017). *Disease as Scientific and as Value-Laden Concept.* T. Schramme, S. Edwards (eds.), Handbook of the Philosophy of Medicine, Springer Science+Business Media Dordrecht 2017, p. 45–62.

Laing, R. D. (1960). *The Divided Self. An Existential Study in Sanity and Madness*. Harmondsworth: Penguin.

Laing, R. D (1967). *The Politics of Experience*. London: Penguin Books.

Margolis, Eric & Laurence, Stephen (2014). "*Concepts*", The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/spr2014/entries/concepts/.

Millon, T. (1975). *Reflections on Rosenhan´s "On Being sane in Insane Places"*. Journal of Abnormal Psychology 1975, Vol. 84, No. 5, 456–461.

Murphy, D. (2015). "*Concepts of Disease and Health*", The Stanford Encyclopedia of Philosophy (Spring 2015 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/spr2015/entries/health-disease/.

Murray, G. W. (2011). *A Critical Introduction to DSM*. Psychology Research Progress. Nova Science Publishers, Inc. New York.

Nordenfelt, L. (1993). *Quality of Life, Health and Happiness*. Ashgate Publishing, Aldershot.

Nordenfelt, L. (2017). *On Concepts of Positive Health*. T. Schramme, S. Edwards (eds.), Handbook of the Philosophy of Medicine, Springer Science+Business Media Dordrecht 2017, p. 29–41.

Oldham, P. D., Pickering, G., Fraser Roberts, J.A. & Sowry, G. S. C. (1960). *The Nature of Essential Hypertension*. Lancet, i, 1085–1093.

Reznek L (1987). *The Nature of Disease*. Routledge, London.

Sadler, J. Z & Agich, G. J. (1995). *Diseases, Functions, Values and Psychiatric Classification*. Philosophy, Psychiatry & Psychology, Volume 2, Number 3, Septempber 1995, p. 219–231.

Scadding, J. G. (1959). *Principles of Definition in Medicine*. The Lancet, February 1959, 323–325.

Scadding, J. G. (1963). *Meaning of Diagnostic Terms in Broncho-Pulmonary Disease*. British Medical Journal, ii, 1425–1430.

Scadding, J. G. (1967). *Diagnosis: the Clinician and the Computer*. Lancet, ii, 877–882.

Scheff, T. J. (1963). *The Role of the Mentally Ill and the Dynamics of Mental Disorder: a Research Framework*. Sociometry, 26, 436–453.

Schwartz, P. (2007). *Defining dysfunction: Natural Selection, Design, and Drawing a Line*. Philos Science 74, p. 364–385.

Shandera, A.L., Berry D.T.R., Clark J.A., Schipper, L.J., Graue L.O., Harp J.P. (2010). *Detection of malingered mental retardation*. Psychological Assessment, Vol 22(1), Mar 2010, 50–56.

Shorter, E. (2005). *A Historical Dictionary of Psychiatry*. Oxford University Press. New York

Silverstein, C. (2008). *The Implications of Removing Homosexuality from the DSM as a Mental Disorder*. Arch Sex Behav (2009) 38:161–163, published online: 11 November 2008.

Spitzer, R. L. (1975). *On Pseudoscience in Science, Logic in Remission, and Psychiatric Diagnosis: A Critique of Rosenhan's "On Being Sane in Insane Places"*. Journal of Abnormal Psychology 1975, Vol. 84, No. 5, 442–452.

Spitzer, R. L., Endicott J. & Franchi J-A. M. (1978). *Medical and Mental Disorder: Proposed Definition and Criteria*. Originally published in Critical issues in psychiatric diagnosis, edited by Robert L. Spitzer and Donald F. Klein. Raven Press, New York, 1978. Available online 9 August 2018.

Spitzer, R. L. (2001). *Values and Assumptions in the Development of DSM-III and DSM-III-R: An Insider's Perspective and a Belated Response to Sadler, Hulgus, and Agich's "On Values in Recent American Psychiatric Classification*. The Journal of Nervous and Mental Disease. Vol. 180, No 6., 351–359. USA.

Stempsey, W. (2000). *A Pathological View of Disease*. Theoretical Medicine and Bioethics, 21(4), p. 321–330.

Szasz, T. S. (1960). *The Myth of Mental Illness*. American Psychologist , 15 , 113–18.

Vehmas, S., Mäkelä P. (2008a). *A realist account of the ontology of impairment*. Journal of Medical Ethics, 34(2), p. 93.

Vehmas, S., Mäkelä P. (2008b). *The ontology of disability and impairment: A discussion of the natural and social features*.

Wakefield, J. C. (1992). *The Concept of Mental Disorder: On the Boundary Between Biological Facts and Social Values*. American Psychologist, 47, 373–388.

Wakefield, J. C. (1995). *Dysfunction as a Value-Free Concept: A Reply to Sadler and Agich*. Philosophy, Psychiatry & Psychology, Volume 2, Number 3, Septempber 1995, p. 233–246.

Wakefield, J. C. (2007). *The concept of mental disorder: Diagnostic implications of the harmful dysfunction analysis*. World Psychiatry, *6*(3), p. 149–156.

Wakefield, J. C. (2010). *False positives in psychiatric diagnosis: Implications for human freedom*.Theoretical Medicine and Bioethics, 31(1), p. 5-17.

Wakefield, J. C. (2011). *Placing symptoms in context: The role of contextual criteria in reducing false positives in Diagnostic and Statistical Manual of Mental Disorders diagnoses*. Comprehensive Psychiatry, *53*(2).

Wakefield, J. C. (2014). *The Biostatistical Theory Versus the Harmful Dysfunction Analysis, Part 1: Is Part-Dysfunction a Sufficient Condition for Medical Disorder?* Journal of Medicine and Philosophy, 39: 648–682, 2014.

Wakefield, J. C. (2016). *Diagnostic Issues and Controversies in DSM-5: Return of the False Positives Problem*. Annual Review of Clinical Psychology, *12*(1), p. 105-132.

Wakefield, J. C. (2017). *Mental Disorders as Genuine Medical Conditions*. In Schramme, T. & Edward, S. (2017). Handbook of the philosophy of medicine. Dordrecht: Springer Netherlands. p. 65–82.

World Health Organization (WHO), (1948). *WHO definition of Health* in Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19–22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.

Xhignesse P. & Krzesinski J.M. (2018). *New 2018 European guidelines for the management of hypertension and comparison with the 2017 American guidelines*. Revue Medicale de Liege 01 Nov 2018, 73(11): 583–591.

Zachar, P. & Kendler, K. S. (2007). *Psychiatric Disorders: A Conceptual Taxonomy*. American Journal of Psychiatry. 164: 557–565.