



# What is mental health and disorder? Philosophical implications from lay judgments

Somogy Varga<sup>1,2</sup> · Andrew J. Latham<sup>1</sup>

Received: 27 October 2023 / Accepted: 5 March 2024  
© The Author(s) 2024

## Abstract

How do people understand the concepts of mental health and disorder? The objective of this paper is to examine the impact of several factors on people's judgments about whether a condition constitutes a mental disorder or a healthy state. Specifically, this study examines the impact of the source of the condition, its outcome, individual valuation (i.e., the value the individual attaches to the condition), and group valuation (i.e., the value the relevant group attaches to the condition). While we find that people's health and disorder judgments are driven by perceived dysfunction, we also find that health and disorder judgments are impacted differently by these factors. Health judgements are impacted by outcome and individual valuation, and disorder judgments are impacted by condition source. These results suggest that the folk concept of mental health is positive (i.e., mental health is more than the absence of mental disorder) and normativist (i.e., value judgments play a significant role in determining whether a condition counts as healthy), while the concept of mental disorder aligns with a naturalist perspective, at least to the extent that dysfunction plays an important role in categorizing a condition as a disorder. However, our finding that people's dysfunction judgments are influenced by individual valuation and outcomes poses a strong challenge to naturalist accounts.

**Keywords** Health · Disease · Experimental philosophy · Philosophy of medicine · Dysfunction

---

✉ Somogy Varga  
varga@cas.au.dk

<sup>1</sup> Department of Philosophy and History of Ideas, Aarhus University, Jens C. Skousvej 7, Aarhus 8000, Denmark

<sup>2</sup> The Centre for Philosophy of Epidemiology, Medicine and Public Health, University of Johannesburg, Johannesburg, South Africa

How health and disease are characterized has important implications for medical research, public health measures, and clinical decision-making. As a result, systematic reflection on the nature of these concepts has garnered considerable and ongoing interest within philosophy of medicine (see e.g., Giroux, 2016; Radden, 2019; Reiss & Ankeny, 2016). Despite the divergence of views on several key issues among philosophers, there are two issues that warrant particular attention. We will refer to these issues as the *evaluative issue* and the *relational issue* (for reviews, see Kingma, 2019; Murphy, 2021).

The *evaluative issue* is whether health and disease are evaluative or non-evaluative concepts. There are three main positions: *naturalism*, which claims that health and disease are non-evaluative concepts (e.g., Boorse, 1977, 1997, 2014), *normativism*, which claims that health and disease inevitably involve evaluative judgments (e.g., Cooper, 2002, 2005; Nordenfelt, 1995; Glacking, 2019; Fulford, 1989), and *hybridism*, which combines descriptive and evaluative elements (e.g., Wakefield, 1992, 2014; Wakefield & Conrad, 2020). The *relational issue* is about how health and disease are related to each other. *Negativism* claims that health is merely the absence of disease (e.g., Boorse, 1977; Wakefield, 2014) whereas *positivism* holds that health is the presence of some positive state or ability (e.g., Nordenfelt, 2017; Venkatapuram, 2013; Wren-Lewis & Alexandrova, 2021).

Despite what is at stake, many researchers now argue that the philosophical debates concerning the concepts of health and disease have hit a standstill, with conceptual analysis stalling between conflicting intuitions (Schwartz, 2017; Lemoine, 2013; Sholl, 2015; Fuller, 2018). There is growing sentiment that conventional conceptual analysis alone cannot propel philosophical debates forward, prompting some to argue in favor of incorporating different methods, including empirical ones, such as those of *experimental philosophy* (De Block & Hens, 2021; Griffiths & Stotz, 2008). Grasping people's understanding of the concepts of health and disease via empirical methods might carry important implications for philosophical debates, as many think that definitions and conceptual analyses should align with common sense judgments when possible.

To date, there have been several studies that have examined the concept of mental disorder, though not with the main aim of contributing to philosophical debates related to the evaluative and relational issue (Kirk et al., 1999; Béghin & Faucher, 2023; Wakefield et al., 2006; Wakefield, 2021; Tse & Haslam, 2023). Instead, their focus has been examining how closely the lay judgments align with the definition of mental disorder described in the DSM-5. Here we will highlight three recent studies that have directly targeted the evaluative and relational issue (Machery, 2023; Varga and Latham, forthcoming; Varga, Latham, and Machery, forthcoming). These studies utilized a *contrastive vignette technique* to investigate how individuals understand and categorize health and disease, what factors influence their decisions to label a condition as health or disease, and how these judgments vary across different demographic groups.

Speaking to the relational issue, one study has found that most lay people, as well as medical students, deploy a positive concept of health, in which health is more than the absence of disease (Varga and Latham, forthcoming). Speaking to the evaluative issue, Machery (2023) examined how people's disease judgments are influenced by

whether the condition is *typical*, involves *dysfunction*, and is *disvalued by the group*. Machery's findings tentatively indicated that the folk concept of disease is naturalistic (i.e., value judgments do not matter for whether a condition is a disease). Finally, with respect to both issues, another study examined the effect of *typicality*, *dysfunction*, *individual valuation*, and *group valuation* on both people's health and disease judgments (Varga, Latham, and Machery, [forthcoming](#)). Supporting naturalism, only dysfunction was found to have a significant effect on health or disease judgments: typicality, individual and group valuation did not appear to play any role in determining whether a condition counts as a disease or someone is healthy.<sup>1</sup>

While these studies provide some empirical traction on these issues, they possess limitations which mean that caution is warranted. First, these studies only considered physical conditions, and it is very plausible that these results would vary significantly if mental conditions were being evaluated. Second, while it seems that people's judgments are only sensitive to dysfunction, very few participants thought that the conditions being evaluated in two studies (i.e., purple eyes leading to color blindness) were a disease. This indicates the existence of important factors still unaccounted for in our understanding of health and disease.

This study extends the findings of the previous research. First, we focus on mental health and mental disorder, noting that while both "disease" and "disorder" are standardly comprehended as involving deviations from functional norms, in psychiatry, conditions are standardly referred to as "disorders" (e.g., Obsessive-Compulsive Disorder) rather than "diseases".<sup>2</sup> Second, we ask participants to evaluate a mental condition that is clearly atypical and involves a dysfunction.<sup>3</sup> While previous studies did not find any effect of individual and group valuation in the case of physical conditions, it is possible that they might influence people's health and disease judgments regarding a mental condition. Third, in addition to individual and group valuation, we also investigate the impact of two new factors: *outcome* (i.e., whether the condi-

<sup>1</sup> In what follows we will be using the terms "lay" and "folk" interchangeably, as it is done in some of the literature (e.g., Machery, 2023). While making a distinction using these terms could be beneficial in some contexts, e.g., you might want to distinguish between folk concepts and lay understandings of scientific concepts, for our purposes nothing hangs on such a distinction.

<sup>2</sup> "Disorder" is usually used to describe conditions with a complex array of symptoms and causes, often involving a combination of physical, mental, and behavioral aspects. In the field of mental health, the use of "disorder" instead of "disease" can reflect the complex nature of these conditions (often involving a mix of biological, psychological, and social factors) and may not have a clear-cut pathophysiology in the same way as many somatic diseases. While we focused on "disorder", we note that it is sometimes used interchangeably with "illness" in the literature. There are, however, good reasons to keep them separate (see e.g., Wilkinson, 2023). Besides differences being reflected in linguistic intuitions, some have argued that illnesses are particulars whereas diseases are universals (Boorse, 1977), that illness is personal level whereas disease and disorder are sub-personal level (Wilkinson, 2023, ch. 2), and that disorder seems closer in semantic space to illness than disease (Wilkinson, 2023, 17). It is possible that laypeople might understand "illness" and "sickness" in a normative manner, as such future research should move beyond considering just disease and disorder and look at illness and sickness as well. We thank an anonymous referee for highlighting this issue.

<sup>3</sup> While "dysfunction" might be interpreted as having both a biological and practical sense, our focus, consistent with the relevant literature, is primarily on biological/psychological interpretations of these terms. We note, however, that on some accounts, the practical sense of dysfunction can be considered an instance of dysfunction in the biological sense (see, for example, Matthewson & Griffiths, 2017). Nothing in our study directly speaks to this issue.

tion results in a beneficial or detrimental effect for the individual) and *source* (i.e., whether the condition is caused by genetic factors or upbringing). These factors were not only chosen to improve our understanding of how people conceptualize health and disease, but also yield insights relevant for philosophical debates. The findings offer insights into how lay views match-up to philosophical views and carry implications extending beyond philosophical discussions, impacting areas like public health initiatives and clinical psychiatry.

The plan for the paper is as follows. In Sect. 1, we provide a brief overview of the existing philosophical and empirical literature in this field. Then, in Sect. 2, we describe the experimental materials, methods, and hypotheses that guided our research. To provide access to our supplementary data, an [appendix](#) has been included. In Sect. 3, we present the results of the study. Finally, in Sects. 4 and 5, we discuss our findings, their implications for philosophical debates, and describe some limitations of our research.

## 1 Vignette-based experimental design

Studies on health and disease in medical sociology, anthropology, and psychology have typically aimed to establish a connection between individuals' beliefs and attitudes concerning a specific chronic medical condition, such as diabetes, and their corresponding behaviors, like dietary practices (Hughner & Kleine, 2004). Very few studies have explored lay conceptions of health and disease. Utilizing different kinds of surveys and unstructured, in-depth interviews, these studies have most frequently identified several major “themes” in people’s conceptualization of health and disease (e.g., health as the absence of illness, as a capacity, as equilibrium), including the presence of cross-cultural differences (e.g., Herzlich, 1973; Weller, 1984; Williams, 1990; Blaxter, 1990; Jensen & Allen, 1994; Hughner & Kleine, 2004; Bishop & Yardley, 2010). These studies typically ask participants to define their concepts, articulate their understanding, or describe a healthy individual they know (e.g., Blaxter, 2010). While such approaches are well-suited to identifying broad “themes” and rich accounts of the narratives that surround health and disease, they are not well-suited to determining how such “themes” intersect, or what happens when they diverge or conflict. For instance, while there is good evidence that lay people are negativists and define health as the absence of disease (Calnan, 1987; McKague & Verhoef, 2003), there is some equally good evidence that they conceptualize health as more than the absence of disease, such as the ability to function according to one’s own expectations (McKague & Verhoef, 2003) or fulfill social roles (Blaxter, 1990; for a review, see Hughner & Kleine, 2004).

Our approach does not seek to identify “themes”, and, consequently, does not ask participants to define, describe or articulate the contents of their concepts of health and disease. Instead, we ask them to *deploy* these concepts by making judgments about various scenarios. By examining patterns of judgments to systematically varied scenarios, we can gather *defeasible* evidence about the content of people’s concepts, even if that content is largely implicit and opaque to people. Our approach thus is

distinct from the aforementioned research in the social sciences, and from orthodox work in philosophy of medicine, which typically relies on conceptual analysis alone.

More specifically, the current study employs a *vignette-based methodology*. Vignettes describing carefully crafted scenarios are presented to participants who are then asked to respond. Using vignettes allows for the manipulation of certain factors while controlling others, making it possible to investigate how judgments are affected by factors that might be difficult to tease apart in real-life scenarios. Vignette-based designs typically consist of controlled factors (which remain constant across vignettes) and experimental factors (which are manipulated across vignettes), allowing for the assessment of their impact on dependent variables (Evans et al., 2015). By comparing people's responses between different vignettes researchers gain evidence about the content of people's concepts.

It is important to note that in the medical and health psychology literature, vignettes have been occasionally used to identify factors that influence medical decisions and variations in healthcare practices (e.g., Payton & Gould, 2023; Bachmann et al., 2008). In addition, "anchoring vignettes" have been used to improve the between-group comparability of self-assessed health surveys (Grol-Prokopczyk et al., 2011). While important work, our study diverges from these by using vignettes to explore the influence of factors on people's evaluations of health and disease.

## 2 Background and cues

The present study involves vignettes in which the person we describe, Katie, is unable to make slow, methodical, deliberative decisions. Her condition is described as atypical (i.e., not possessed by most people) and dysfunctional (i.e., interferes with normal functioning). Many theorists judge that a condition being atypical and dysfunctional is *necessary* (or even *sufficient*) to be pathological. In Christopher Boorse's Biostatistical Theory (BST), dysfunction is both a necessary and sufficient condition for disease or mental disorder, while in Jerome Wakefield's Harmful Dysfunction Analysis (HDA) (1992, 2014) dysfunction is a necessary condition. Moreover, dysfunction must be tightly linked to atypicality. For instance, in BST, the function of a trait in a reference class is its statistically typical contribution to survival and reproduction, and a pathological condition is species-subnormal part-function (Boorse, 1977, 1997, 2014). Thus, whether something counts as subnormal function depends on levels of functioning in the reference class, and conditions that are typical in a reference class will not count as pathological, even if they result in a decrease in survival and reproduction (see Schwartz, 2007, for a critique).

In this study, we systematically compared the effect of four factors—source, outcome, individual valuation, and group valuation—on people's health and disorder judgments. Each factor plays an important role in the evaluative and relational issues described earlier in the paper.

To address the *relational issue*, we asked people to make both health and disorder judgments, to see whether they would be differentially affected by these factors. If they are, then this would challenge both the BST and the HDA, which embrace negativism and comprehend health as merely the absence of disease and not the pres-

ence of some positive state (e.g., Boorse, 1977, 1997, 2014; Wakefield, 1992, 2005, 2014). While we anticipate that evaluations of health and disorder will diverge, and so might align with positivism, our approach would also enable us to explore to what extent individuals understand health as a favorable condition or capability, possibly associated with well-being. Both the World Health Organization and several philosophers have tied health to the possession of certain skills or abilities that are essential for achieving well-being. For example, Lennart Nordenfelt argues that the second-order abilities that characterize health are those that are necessary for pursuing “vital goals,” where a vital goal for a person is something that is either part of or necessary for the person to achieve a minimum level of happiness or well-being. Wren-Lewis and Alexandrova (2021, 696) present an account of mental health grounded in well-being, proposing a definition of mental health as “the capacities of each and all of us to feel, think, and act in ways that enable us to value and engage in life.” Sridhar Venkatapuram (2011, 2013) argues that health is a necessary precondition for well-being and sees health as the ability to have abilities that are objective and universal conditions for basic well-being. Finally, Graham (2010) argues mental disorder involves an impairment in a fundamental psychological ability required to lead any kind of a “decent or personally satisfying life” (Graham, 2010, 131–132). Overall, while we expect that the patterns of judgments we find to be of relevance for the relational issue (likely aligning with positivism), our approach will also enable us to address accounts that link health to well-being or to the possession of skills or abilities that are essential for well-being.

To address the *evaluative issue* in philosophical debates, we hypothesized that the key factors would be individual valuation, group valuation, and outcome. Valuation (individual and group) is of key interest in the debate between naturalism, normativism and hybridism. While naturalists hold that health and disease are value-neutral properties independent of value judgments, normativist (e.g., Nordenfelt, 2007; Cooper, 2002) and hybrid perspectives on health and disease (e.g., Wakefield, 1992) maintain that the concepts of health and disease are value-laden, reflecting what is deemed valuable or otherwise. Normativists argue that whether something counts as a pathological condition will depend on evaluative judgments about it being undesirable or desirable (Cooper, 2002, 2005), whereas the HDA holds that harm is a prerequisite for a condition to be classified as a disease or disorder.

Mirroring a division found in the philosophical literature, the distinction we introduce between individual and group valuation will allow us to address specific accounts more directly. Some philosophers think that what matters is individual valuation. For instance, Nordenfelt defines health in relation to the individual’s ability to successfully pursue “vital goals” leaving open the possibility that agents may have different vital goals, depending on what they value in life. One consequence of this view is that the lack of a certain ability might undermine the health of one person with a particular set of values but not another (see Cooper, 2002, 2005; Wren-Lewis & Alexandrova, 2021). In contrast, others think that what matters is group valuation. For example, according to the HDA, what counts as a *harmful* condition depends not on individual valuation, but social norms that stem from the values of the culture that the individual is a member of. One consequence of this view is that a particular trait could be viewed as a disorder in one culture but not in another. Other accounts deny

that health and well-being are agent-dependent. For instance, Sridhar Venkatapuram (2011, 2013) argues that health is a necessary precondition for well-being and sees health as the ability to have abilities that are objective and universal conditions for basic well-being. Thus, for Venkatapuram, a person's health (and well-being) can be compromised even if the person does not disvalue the condition and does not consider herself unhappy. Something similar holds for Graham's account, which comprehends the relevant abilities (e.g., the ability to understand oneself and the world, to take responsibility for oneself and make decisions) as Rawlsian primary goods that everyone would prefer in the original position, because these goods are compatible with all ideas of the good life (Graham, 2010, 147–149).

The factor outcome introduces an aspect that is highly relevant for both the *evaluative issue* and the *relational issue*. While the factors individual and group valuation are concerned with evaluations of the condition, the factor outcome is concerned with the outcomes of that condition. While these evaluations are no doubt tightly linked, they nevertheless can come apart. For example, I might have a negative attitude towards a condition that I possess, even if that condition results in positive outcomes for me. Speaking first to the evaluative issue, the outcome of a condition (i.e., whether it leads to positive or negative results for the person with the condition) does not matter for whether or not the condition warrants the label "disorder." Instead, for naturalists, what matters is the existence of a dysfunction, a deviation or failure in normal physiological or psychological functioning such as an inability to make decisions in a psychologically flexible manner, which is generally expected in healthy individuals. Outcome does not directly matter because it could be attributable to random chance and is not intrinsically tied to the nature of the condition itself. In contrast, normativists and hybridists could allow outcome to matter for health and disorder judgments, but only via individual or group valuation. This means that the outcome of a condition could influence whether it is considered a disorder, but only if that outcome is deemed undesirable or harmful by the standards of the individual or the relevant group. To explore this further, we defined positive outcomes as those that support the individual in realizing her vision of a good life, enhancing her well-being, and promoting long-term happiness. Conversely, negative outcomes are those that obstruct or hinder her in achieving these aspirations. By distinguishing between positive and negative outcomes in this way, we can examine the interplay between the nature of the condition, its outcomes, and different philosophical perspectives on what constitutes a disorder.

The factor outcome is also relevant for the *relational issue*. Expanding on our exploration of positivism, we hypothesized that adding outcome to the design of our study would provide us with insights into how the abilities associated with health and disorder are conceptualized. Positivist accounts will agree that something like being in important relationships or making reflected decisions are vital goals and that the compromised or absent ability to achieve them counts as a disorder. But what if one succeeds in attaining their vital goal without possessing the relevant ability? On some views like Graham's, the condition should still count as a disorder. This, however, is not what we predicted.

Finally, as with outcome, orthodoxy in the philosophical debates holds that the source of a condition is not relevant to whether a condition is a disorder. This also

aligns with some recent empirical work. For instance, Machery (2023) tested to what extent different physical sources of a condition (i.e., genetic mutation, bacterium, nuclear power plant) impacted people's disease judgments. Machery's investigations were motivated by early anthropological accounts of disease that highlighted the nature of a condition's cause as being relevant to people's disease judgements (e.g., Clements, 1932; Rogers, 1944; Young, 1978; Foster, 1976). Ultimately, Machery did not find any effect of the different physical sources. Nonetheless, there are good reasons to think that results could be different when comparing physical versus social causes, especially when evaluating mental disorders. For instance, Machery (2023) maintains that our disease concept is part of folk biology, but it is plausible that (certain) mental disorders might instead be part of folk psychology. Furthermore, while lay beliefs about the origins of disease typically align with knowledge derived from scientific research, this is often not the case with mental disorders where people's etiological beliefs appear to matter (Troisi & Dieguez, 2022). Interventions on physical disorders most often proceed via underlying physical mechanisms, but interventions on mental disorders can also proceed via psychosocial mechanisms. As a result, we predicted that the source of a condition might influence whether it is considered a disorder. Specifically, we predicted that conditions that result from a biological source would be more likely to be viewed as disordered. In contrast, conditions that result from social factors would be less likely to be viewed as disordered. Instead, they might be viewed as something else, for example, a "lifestyle problem".

### 3 Methods and results

The study was pre-registered at <https://osf.io/fy23w>. 400 people were recruited online using Prolific. 53 were excluded from the analyses for failing to respond to all the questions or answer all the attention and comprehension checks correctly. The final sample consisted of 347 participants (173 female, 8 trans/non-binary, aged 19–79;  $M=38.75$ ,  $SD=13.17$ ). Ethics approval for the study was obtained from the Aarhus University Human Ethics Committee.<sup>4</sup>

The study was a 2 (source: genetic vs. upbringing)  $\times$  2 (outcome: positive vs. negative)  $\times$  2 (individual: positive vs. negative)  $\times$  2 (group: positive vs. negative) between-subjects design. Participants were randomly assigned to one of 16 conditions. The vignettes of the study read as follows (the vignettes vary across conditions between brackets):

Human decision-making has evolved to include two distinct mental systems, known as System 1 and System 2. System 1 is characterized by unconscious, rapid and intuitive decision-making, and is typically used to make quick and routine decisions. Since System 1 decision-making operates outside of our

---

<sup>4</sup> We used G\*Power to conduct a power analysis. Our goal was to obtain 0.9 power to detect a medium effect size at 0.005 alpha error probability (Benjamin et al., 2018). We calculated that a sample size of 272 participants would be required. To be safe, we increased the sample by 45% to account for attention and comprehension check failures. This gave us a total sample size of 395, which was rounded up to 400.



conscious awareness, we have limited conscious control over these intuitive choices, and we typically do not know why we make them. In contrast, System 2 is characterized by conscious, slow, methodical, and deliberate decision-making, and is typically used to make complex and important decisions. Since System 2 decisions rely on conscious reflection, we have conscious control over them and we typically know why we make them. Most experts believe that both systems are necessary, no matter what conceptions of the good life, well-being, or long-term happiness we aim to attain.

Katie is in all respects an ordinary woman, but she possesses a genetic mutation that determines that no matter what kind of decision she has to make, Katie always uses System 1 and makes very fast and extremely intuitive decisions. Even if she tries to use System 2 to make slow, considered decisions, she almost never succeeds. The fast System 1 always kicks in and makes decisions before she has a chance to engage her more deliberate System 2. (Katie is in all respects an ordinary woman, but her unique upbringing and education determine that no matter what kind of decision she has to make, Katie always uses System 1 and makes very fast and extremely intuitive decisions. Her parents and mentors emphasized the importance of taking action quickly and she was constantly exposed to risky situations where speed and efficiency were very important. Even if she tries to use System 2 to make slow, considered decisions, she almost never succeeds. The fast System 1 always kicks in and makes decisions before she has a chance to engage her more deliberate System 2.)

Katie's very fast and extremely intuitive decisions almost always bring about positive outcomes that help her achieve her conception of the good life, well-being, and long-term happiness. As a result, Katie is able to excel in her career, manage her finances effectively, and establish lasting personal relationships. (Katie's very fast and extremely intuitive decisions almost always bring about negative outcomes that hinder her in achieving her conception of the good life, well-being, and long-term happiness. As a result, Katie is unable to keep a job, manage her finances effectively, and establish lasting personal relationships.)

Katie does not mind that she frequently remains unaware of the reasons behind her choices, and she does not perceive the way she makes decisions as harmful or negative for her life. She feels content with it, it brings her a sense of calm and security, and she can not imagine herself thinking differently and judging differently than the way she in fact does now. (Katie minds that she frequently remains unaware of the reasons behind her choices, and perceives the way she makes decisions as harmful and negative for her life. She feels really unhappy with it, it causes her profound insecurity and distress, and she wishes that she could change the way she thinks so that she could succeed in slowing and reflecting before deciding.)

According to the prevailing social norms and values of the society in which Katie lives, the way she makes decisions can be valuable and have a positive impact on achieving a good life, well-being, and long-term happiness. (According to the prevailing social norms and values of the society in which Katie lives, the way she makes decisions is disvalued and considered to have a negative impact on achieving a good life, well-being, and long-term happiness.)

Table 1 below shows all 16 possible conditions, of which participants saw and responded to one. Following the vignette participants were asked: “In this scenario, Katie is healthy.”; “In this scenario, Katie has a mental disorder.” To which participants could indicate their level of agreement on a 7-point Likert scale that ranged between “Strongly disagree” and “Strongly agree”. We also asked the following comprehension check questions: (A) “In this scenario, System 1 is associated with unconscious, rapid, and intuitive decision-making.”; (B) “In this scenario, System 2 is associated with conscious, slow, methodical, and deliberate decision-making.”; (C) “In this scenario, Katie almost always uses System 2 to make decisions.” Once again participants could indicate their level of agreement on a 7-point Likert scale that ranged between “Strongly disagree” and “Strongly agree”. Participants who failed to agree to (A) and (B) and disagree with (C) were excluded from the analyses.

While our scenario described a condition that many theorists would judge to be a dysfunction, this does not mean that it is something that most lay people would consider to be a dysfunction. To account for this fact, we asked participants “In this scenario, Katie’s decision making is dysfunctional”. To which participants could indi-

**Table 1** *Experimental Conditions*

Condition
1 Genetic source/Positive outcome/Positive individual valuation/ Positive group valuation
2 Genetic source/Positive outcome/Positive individual valuation/ Negative group valuation
3 Genetic source/Positive outcome/Negative individual valuation/ Positive group valuation
4 Genetic source/Positive outcome/Negative individual valuation/ Negative group valuation
5 Genetic source/Negative outcome/Positive individual valuation/ Positive group valuation
6 Genetic source/Negative outcome/Positive individual valuation/ Negative group valuation
7 Genetic source/Negative outcome/Negative individual valuation/ Positive group valuation
8 Genetic source/Negative outcome/Negative individual valuation/ Negative group valuation
9 Upbringing source/Positive outcome/Positive individual valuation/ Positive group valuation
10 Upbringing source/Positive outcome/Positive individual valuation/ Negative group valuation
11 Upbringing source/Positive outcome/Negative individual valuation/ Positive group valuation
12 Upbringing source/Positive outcome/Negative individual valuation/ Negative group valuation
13 Upbringing source/Negative outcome/Positive individual valuation/ Positive group valuation
14 Upbringing source/Negative outcome/Positive individual valuation/ Negative group valuation
15 Upbringing source/Negative outcome/Negative individual valuation/ Positive group valuation
16 Upbringing source/Negative outcome/Negative individual valuation/ Negative group valuation

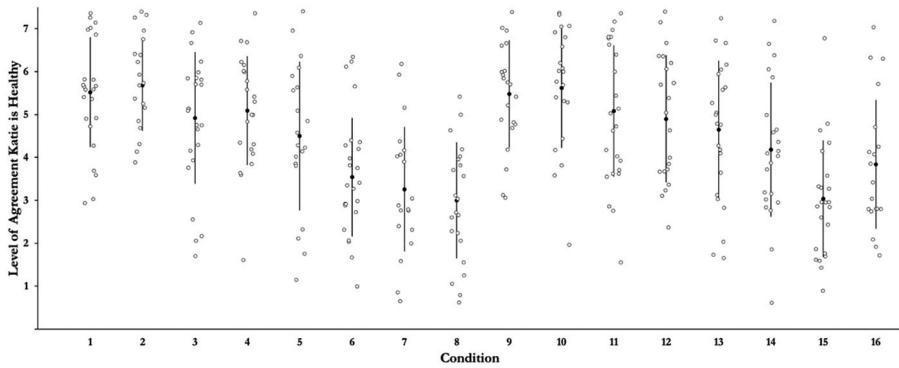
cate their level of agreement on a 7-point Likert scale that ranged between “Strongly disagree” and “Strongly agree”. Similarly, it is also possible that the extent to which participants judge that a scenario could actually take place actually could impact their health and disorder judgments too. To account for this, we also asked “How likely do you think it is that the scenario you were asked to read could actually take place in the world?” To which participants could indicate the level of likelihood on a 7-point Likert scale that ranged between “Incredibly unlikely” and “Incredibly likely”.

We also asked a number of further questions regarding related phenomena to health and disorder. We did not have any specific hypotheses regarding how the different factors in the scenario might impact people’s judgments about these, but we were interested in exploring what people might say. We asked: “In this scenario, Katie is morally responsible for her decisions.”; “In this scenario, Katie’s condition impacts her ability to achieve her goals.”; “In this scenario, Katie is in control of her decisions.”; “In this scenario, Katie is the author of her decisions.” To which participants could indicate their level of agreement on a 7-point Likert scale that ranged between “Strongly disagree” and “Strongly agree”. Finally, we asked participants “Please rate the level of well-being that you perceive in Katie.” To which participants could indicate perceived level of well-being on a 7-point Likert scale that ranged between “Low well-being” and “High well-being”. Results for these exploratory questions can be found in [Appendix A](#) of this paper.

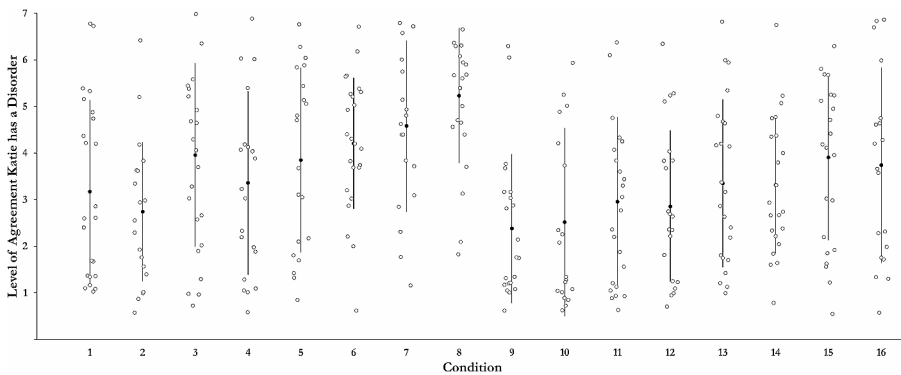
We first examined participants’ judgements across conditions using a MANOVA. Source, outcome, individual valuation, and group valuation were entered as between-subjects factors. Gender, age, political ideology, and vignette possibility were entered into the analysis as covariates. The same factors and covariates were entered into each of the analyses which follow, unless otherwise stated. The results of the analysis revealed a significant main effect of source,  $\Lambda=0.900$ ,  $F(8, 319)=4.451$ ,  $p<.001$ ,  $\eta_p^2=0.100$ , outcome,  $\Lambda=0.421$ ,  $F(8, 319)=54.915$ ,  $p<.001$ ,  $\eta_p^2=0.579$ , and individual valuation,  $\Lambda=0.793$ ,  $F(8, 319)=10.421$ ,  $p<.001$ ,  $\eta_p^2=0.207$ . There was also a significant effect of the covariate possibility,  $\Lambda=0.851$ ,  $F(8, 319)=7.006$ ,  $p<.001$ ,  $\eta_p^2=0.149$ . No other significant effects were observed. Next, we report results of separate ANOVAs that show the effects of these factors on participants’ judgments.

Figure 1 below shows the descriptive results for participant’s health judgments. The ANOVA revealed a significant main effect of outcome,  $F(1,326)=99.702$ ,  $p<.001$ ,  $\eta_p^2=0.234$ , and individual valuation,  $F(1,326)=25.601$ ,  $p<.001$ ,  $\eta_p^2=0.073$ . The main effect of outcome was that participant’s health judgments were significantly lower in the negative cases ( $M=3.74$ ,  $SD=1.44$ ) than in the positive cases ( $M=5.30$ ,  $SD=1.44$ ). The main effect of individual valuation was that participant’s health judgments were significantly lower in the negative cases ( $M=4.13$ ,  $SD=1.42$ ) than in the positive cases ( $M=4.91$ ,  $SD=1.43$ ).

Figure 2 below displays the descriptive results for participant’s disorder judgments. The ANOVA revealed a significant main effect of source,  $F(1, 326)=12.810$ ,  $p<.001$ ,  $\eta_p^2=0.038$ , outcome,  $F(1, 326)=29.360$ ,  $p<.001$ ,  $\eta_p^2=0.083$ , and individual valuation,  $F(1, 326)=11.259$ ,  $p<.001$ ,  $\eta_p^2=0.033$ . The main effect of source was that participant’s disorder judgments were significantly lower in the upbringing cases ( $M=3.17$ ,  $SD=1.77$ ) than in the genetic cases ( $M=3.85$ ,  $SD=1.76$ ). The main effect of outcome was that participant’s disorder judgments were significantly lower in the



**Fig. 1** Jitter plot showing the distribution of participant responses to the question “In this scenario, Katie is Healthy”. Black dots represent the mean response value and error bars show standard deviation

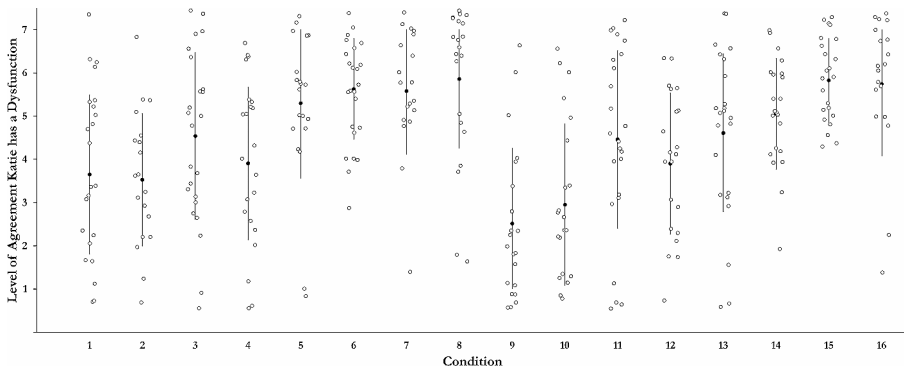


**Fig. 2** Jitter plot showing the distribution of participant responses to the question “In this scenario, Katie has a mental disorder”. Black dots represent the mean response value and error bars show standard deviation

positive cases ( $M=2.99$ ,  $SD=1.77$ ) than in the negative cases ( $M=4.03$ ,  $SD=1.77$ ). The main effect of individual valuation was that participant’s disorder judgments were significantly lower in the positive cases ( $M=3.19$ ,  $SD=1.76$ ) than in the negative cases ( $M=3.83$ ,  $SD=1.75$ ).

Figure 3 below displays the descriptive results for participant’s dysfunction judgments. The result of ANOVA showed a significant main effect of outcome,  $F(1, 326)=93.024$ ,  $p<.001$ ,  $\eta_p^2=0.222$ , and individual valuation,  $F(1, 326)=20.728$ ,  $p<.001$ ,  $\eta_p^2=0.060$ . The main effect of outcome was that participant’s disorder judgments were significantly lower in the positive cases ( $M=3.70$ ,  $SD=1.65$ ) than in the negative cases ( $M=5.43$ ,  $SD=1.65$ ). The main effect of individual valuation was that participant’s disorder judgments were significantly lower in the positive cases ( $M=4.16$ ,  $SD=1.64$ ) than in the negative cases ( $M=4.97$ ,  $SD=1.65$ ).

Given the influence that outcome and individual valuation were observed to have on participant’s dysfunction judgments we were interested in exploring whether the influence that outcome and individual valuation had on people’s health and dis-



**Fig. 3** Jitter plot showing the distribution of participant responses to the question “In this scenario, Katie’s decision making is dysfunctional”. Black dots represent the mean response value and error bars show standard deviation

ease judgements was a direct influence, or an indirect influence via dysfunction. To explore this possibility, we reran both health and disorder ANOVAs with the inclusion of participant’s dysfunction judgments. If outcome and individual valuation only indirectly influence participant’s health and disorder judgments, then the inclusion of participant’s dysfunction judgments should block their previously observed effects.<sup>5</sup>

First, when we reexamined participant’s health judgements, we continued to observe significant main effects of both outcome,  $F(1, 325)=39.929, p<.001, \eta_p^2=0.109$ , and individual valuation,  $F(1, 325)=13.212, p<.001, \eta_p^2=0.039$ , as well as a significant effect of dysfunction,  $F(1, 325)=73.650, p<.001, \eta_p^2=0.112$ . The effect of dysfunction was that higher dysfunction judgments were associated with lower health judgments.

In contrast, when we reexamined participant’s disorder judgments, although we continued to observe a significant main effect of source,  $F(1, 325)=9.551, p=.002, \eta_p^2=0.029$ , as well as a significant effect dysfunction,  $F(1, 325)=127.807, p<.001, \eta_p^2=0.343$ , we failed to observe any significant effect of outcome,  $F(1, 325)=0.094, p=.760, \eta_p^2<0.001$ , or individual valuation,  $F(1, 325)=1.146, p=.285, \eta_p^2=0.004$ . The effect of dysfunction was that higher dysfunction judgments were associated with higher disorder judgments. Thus, it appears that while outcome and individual valuation might have a direct influence on participant’s health judgments, they might only have an indirect influence on participant disorder judgments via dysfunction.

## 4 Discussion

The primary objective of our paper was to investigate the concept of mental disorder and contribute to philosophical debates regarding the evaluative and relational aspects of this issue. In the following subsections, we will address the implications of our findings on these matters.

<sup>5</sup> We would like to thank an anonymous referee for this suggestion.

## 4.1 The evaluative issue

Beginning with the evaluative issue our findings have implications for the debate between naturalism, normativism and hybridism. First, consistent with previous research in experimental philosophy of medicine, our findings suggest that naturalism and hybridism might be correct that dysfunction is central to the lay concept of disorder (e.g., Wakefield, 2021; Béghin & Faucher, 2023). However, in contrast to those earlier results, we found that the factors outcome and individual valuation influence participant's dysfunction judgments. Second, our findings also show that the factors outcome and individual valuation influence participant's health judgments. This suggests that the lay concept of health might not be value-neutral. Let us take a closer look at each of these findings in turn.

One key finding of this study was the significant association between participant's *dysfunction* judgments and both their health and disease judgments. This association is typically taken to be evidence that people might be naturalists or hybridists with respect to health and disease (e.g., Machery, 2023), but that might not be the case. First, as we already noted, participant's health judgments in this study were significantly influenced by the factors outcome and individual valuation; the latter of which, at least, is certainly not value-neutral. Second, whether the association between dysfunction and disorder provides any evidence in favor of naturalism or hybridism ultimately depends on people's understanding of dysfunction. The description of the mental condition that we used in our study was developed to meet a value-neutral description of dysfunction. As such, we did not anticipate that any of the factors that we examined would influence whether participants would consider the condition to be a dysfunction (or not). Surprisingly, however, we found that both the factors outcome and individual valuation influenced participant's dysfunction judgments. As a result, people might not actually be naturalists or hybridists with respect to disorder either. That is because, even if outcome and individual valuation do not directly influence participant's disorder judgments, they can indirectly influence them via their influence on dysfunction.

Of course, the thought that people are neither naturalists nor hybridists with respect to disorder hinges on it being the case that the influence of the factors outcome and individual valuation provide evidence of value-ladenness. First, the influence of the factor individual valuation in our study suggests that what people's dysfunction judgments regarding a condition depend on, at least in part, is the attitude of the person who possesses condition towards it. This result is consistent with those of a recent study performed by Latham and Varga ([forthcoming](#)). They directly examined whether a patient's evaluation of their condition influenced dysfunction judgments and found evidence that they influence participant's judgments in the case of mental conditions, such as the condition in this study, but not physical conditions. This result suggests then that while individual's valuations can influence dysfunction judgments, this influence might *not* generalize across all cases. Developing our understanding of how individual evaluations contribute to people's understanding of dysfunction is an important direction for future research.

Second, whether the influence of the factor outcome suggests that people's dysfunction judgments are value-laden will depend on precisely what is driving the influ-

ence of the factor outcome. In the current study, positive outcomes are characterized in terms of being whatever it is “that helps her [Katie] achieve her conception of the good life, well-being, and long-term happiness”. In contrast, negative outcomes were characterized in terms of being whatever it is “that hinders her [Katie] in achieving her conception of the good life, well-being, and long-term happiness”. While the fact that outcomes are evaluated positively by the agent is made salient by the vignette, it is possible that the outcomes, whatever they are, might also count as positive or negative *despite* how the agent evaluated them. People might think (perhaps tacitly) about positive and negative outcomes in a value-neutral manner. For instance, achieving one’s conception of a good life, well-being, and long-term happiness is a positive outcome because it reliably contributes to survival and reproduction. Conversely, a failure to achieve these things counts as a negative outcome because it reliably hinders survival and reproduction. Of course, both these value-laden and value-neutral senses of positive and negative outcomes are very likely to track close together, and it is possible that both might influence people’s judgments. Further research is required to disentangle these two senses of the factor outcome.

Surprisingly, our results also showed that the factor condition source significantly influenced participant’s disorder judgments, but not their health judgments. Specifically, participants were more likely to judge that the presented condition was a disorder when it had a genetic cause rather than a social cause. Most theorists judge that the source of a condition is irrelevant to whether it is a disorder or not. So why do we find evidence that it impacts lay people’s judgments? One explanation might be that the source of a condition is acting as a proxy for some other *relevant* factor. For instance, Varga, Latham, and Machery ([forthcoming](#)) suggest that dysfunction *magnitude* might be important for people’s disease judgments. People might judge that a condition with a genetic source is more likely to be a disease or disorder than the same condition with a social source because dysfunctions associated with genetic factors are *reliably* more severe than those associated with social conditions. Of course, the influence of condition source might also lend itself to either normativist or hybridist interpretations. Perhaps the reason source is associated with higher disorder judgments is not because genetic sources are reliably associated with higher severity than social sources, but because people *negatively value* genetic sources more than social sources. That said, if condition source is indeed tracking something evaluative, then it is not clear why it did not also affect health judgments. Future research is required to find out what it is about condition source that matters to people making disorder judgments.

Finally, what about the influence of individual valuation and outcome on participant’s health judgments? First, the finding that individual valuation influences health judgments provides support to normativist and hybridists positions that highlight the importance of an individual’s valuation of their condition. But what about those normativist and hybridist accounts which hold that it is the *group’s valuation* that matters (e.g., Wren-Lewis & Alexandrova, 2021; Graham, 2010; Wakefield, 1992, 2014)? For instance, according to the HDA, while harm is a necessary condition for disorder, what counts as a *harmful* condition depends on social norms rooted in the cultural values of the individual’s community. We failed to find any evidence at all that group valuation has any impact on participant’s judgments.

How about the factor outcome? Whether the result of outcome provides any evidence against naturalism about health will depend on, as described above, what it is about outcome that is driving participants responses. For instance, if what is driving the responses are outcomes that are positively valued by the person with the condition, then this would count as an empirical mark against naturalism regarding health. That is because according to naturalism whether the outcomes are positively valued (or not) should not matter to whether the person with the condition is healthy or not. Alternatively, if what is driving the responses are outcomes that are positive despite how the person evaluates them (i.e., contributes positively to survival and reproduction), then such a result would be consistent with naturalism. Once again, it is entirely possible that both sense of outcomes matter for people's judgments, and future research is needed to investigate this possibility.

Interestingly, with respect to the factor outcome, difficulties emerge for those normativists that link health to well-being. On normativist accounts by Nordenfelt, Wren-Lewis, Alexandrova, and Graham, Katie is considered to have a mental disorder, because she lacks a certain capacity necessary for attaining some minimal *objective* well-being and good life, which on these accounts would also mean that she is unhealthy.<sup>6</sup> Bracketing for the moment whether the influence of outcome in our study is value-laden, our findings do not align with these views: in scenarios, in which Katie was the recipient of positive outcomes (defined as those that help her achieve *her* conception of the good life, *subjective* well-being, and long-term happiness), people tended not to judge that she is unhealthy. That is, people appear to be moved by the fact that Katie achieves what she wants by her lights.

These normativists might propose an alternative reading of our results. They could highlight that normativism is correct linking health with well-being, but stress that (a) people may not view the relevant ability in Katie's case as necessary for well-being or (b) they may think that the relevant ability is necessary, but also that Katie actually possesses the ability, given that a positive outcome is achieved. In other words, people may understand ability as *actual* instead of *dispositional*. If ability is interpreted in the actual sense, then what makes something unhealthy is that it *under actual circumstances* diminishes or removes an ability that is somehow crucial for well-being. But, if ability is understood dispositionally, then what makes something unhealthy is that it would have impaired well-being, even though it does not under present circumstances. Unfortunately, our study does not yield direct insights on this matter, underlining the necessity for future research.

## 4.2 The relational issue

Moving on to the *relational issue*. Negativism would predict that people's health and disorder judgments go together, while positivism allows that they can come apart such that people's disorder judgments can be affected by a factor that does not affect health judgments. Our results confirm that health and disorder are related and associ-

<sup>6</sup> This is because she lacks a second-order capacity to achieve the first-order capacity to make System 2 decisions (Nordenfelt), the capacity to engage in life due to an absence of self-understanding and agency (Wren-Lewis and Alexandrova), and an impairment in a basic psychological capacity (Graham).



ated with dysfunction. Specifically, higher judgments that the target person has a dysfunction were associated with higher judgments that the person has a disorder, and lower judgments that the person is healthy. This result is also consistent with previous findings in experimental philosophy of medicine which have found a similar effect of dysfunction on people's health and disease judgments (Machery, 2023; Varga and Latham, forthcoming).

*Prima facie* then our results would appear to support negativism, but such a conclusion would be too hasty. Let us distinguish two different forms of negativism. *Strong* negativism, which is how negativism is standardly characterized in the literature, judges that health is the absence of disease. Thus, if someone judges that someone is healthy, then they should also judge that they do not have a disorder, and similarly, if someone judges that someone has a disorder, then they should also judge that they are not healthy. *Weak* negativism, on the other hand, just holds that there is an association between people's health and disease judgments, such that higher health judgments tend to be associated with lower disorder judgments. Thus, *weak* negativism is consistent with judging that someone is healthy and has a disorder. For instance, having increased credence that someone has a disorder might be associated with having reduced credence that someone is healthy, without then judging that the person is unhealthy. Strong negativism implies weak negativism but not vice versa. The question is which version our results appear to support.<sup>7</sup>

Previously we described studies which found that only dysfunction impacted people's health and disease judgments, decreasing health judgements, and increasing disease judgments (Machery, 2023; Varga, Latham, and Machery, forthcoming). However, overall, people judged that the evaluation target was healthy and did *not* have a disease. Thus, while people's health and disease judgments were coupled together in the manner that negativism would predict, more factors, in addition to dysfunction, seem to be required for people to judge whether an evaluation target has a disorder or is unhealthy.

Our results present preliminary evidence of what some of those factors might be. Interestingly, what we found was that those additional factors were different between health and disorder. Specifically, we found that (a) condition source influences people's disorder judgments but not their health judgments, whereas (b) individual valuation and outcome influence people's health judgments but not their disorder judgments (at least not directly). The fact that people's health and disorder judgments are influenced differently by factors suggests that under certain circumstances they could come apart in a way that would be inconsistent with strong negativism.

---

<sup>7</sup> A small clarification is in order regarding this distinction. Strong negativism asserts that it is impossible, conceptually, for something to count as being both a disease and healthy. Positivism, in contrast, asserts that it is entirely possible, conceptually, for something to count as being both a disease and healthy. What we are calling weak negativism is just the claim that there is *some* association between disease and health (apart from the conceptual incompatibility posited by strong negativism). Of course, positivism is consistent with weak negativism but is also consistent with denying weak negativism. That is because a positivist might judge that there is no association at all between disease and health. While it is doubtful whether anyone holds such a perspective, it nevertheless helpfully illustrates how these views can be differentiated. We would like to thank an anonymous referee for pushing for more clarity on this issue.

It is important to note that the BST and the HDA differ in important respects in how they view negativism. The former (Boorse, 1997, 2014) holds that its analysis only applies to the conception of disorder found in theoretical medicine, whereas the latter (Wakefield, 1992, 2007) maintains that it applies to both medical and lay conceptions. This means that our findings carry more substantial implications for the HDA than for the BST. Nevertheless, even if Boorse is right that health professionals operate with a negative conception and our study is right that lay people do not, this has important implications. If health professionals and patients operate with different concepts, then there could be disagreement regarding when health is decreased due to some pathological condition or when health is restored after a treatment. Of course, they will very likely be aligned in most cases given the role that dysfunction plays in both people's health and disorder judgments, but in cases where they do not, the disparity might lead to miscommunication regarding appropriate care and treatment. Further research is needed to explore whether health professionals operate with a strong negative concept, and under what conditions professional and lay judgments come apart.

## 5 Limitations

While this study contributes to our understanding of people's concepts of health and disorder, along with the factors that influence their health and disorder judgments, it is important to highlight some of the limitations of the employed approach. Considering these limitations in the context of future research can contribute to a more nuanced perspective of this matter.

First, while employing vignettes as a methodological approach offers a controlled way to present scenarios and elicit judgments, the use of hypothetical scenarios can introduce a high level of abstraction that could affect participants' responses compared to real-life contexts. The results of the MANOVA indicated, overall, that vignette likelihood had an impact on people's judgments, and while follow-up tests did not find any effect of this factor on people's health and disorder judgments, the details of the scenario and how closely people perceive them to be like an actual scenario very likely impact people's judgment. With that said, it is worth noting that the ability to manipulate certain factors of interest might be incredibly difficult with certain real-life cases. For instance, imagine a case of depression where the individual and group evaluates the condition positively. While a case where everyone positively evaluates depression, as we understand it, is certainly possible, such a situation is hypothetical and removed from real-life situations.

Essentially, we have shifted the limitation from one part of the vignette to another. Developing an understanding of how people understand health, disease, and disorder, will likely depend on examining people's judgments for both real-life and hypothetical cases. Also, to properly isolate the factors we wanted to study, we used a single vignette about a single character. It is possible that there is something particular about this case that causes people to respond in the way that they are. Future research should check whether the pattern of judgments that we observed in this case generalizes to other scenarios.

Second, the current study investigated health and disorder judgments of English-speaking Americans. There is an open question whether these judgments generalize to other Western societies and then beyond that. This is especially important in the context of examining normativist (and hybridists) positions, where what matters is the valuation of the individual or the group. Between-group variability in valuations will not just impact people's judgments to the scenarios but also how they understand the scenarios themselves. Certain participants might perceive the group valuation described in the scenario as aligning with their own group's values, while other participants might judge that they are in conflict with their own group's values. As a result, the generalizability of these findings to other cultural contexts with potentially different characterizations of health and disorder will be limited. Future studies should aim to include more diverse samples and investigate potential cross-cultural variation.

Third, there is still an open question whether the concept deployed by lay people is continuous with the one deployed by health practitioners and researchers. Some accounts of disorder are explicit that they are accounts of the technical concept deployed by health professionals (i.e., BST), whereas others claim to be continuous with both the technical and lay concept (i.e., HDA). To date, there is at least some evidence suggesting that those training to work in medicine make judgments comparable to those of lay people (Varga and Latham, [forthcoming](#)), but it is an open question whether those currently working in the discipline are similarly moved by these factors. This is not to say that if there are differences then revisions must be made on the part of health professionals, rather an awareness of such differences may be important to satisfying the aims of medicine.

## 6 Conclusion

Our study explored how people understand the concepts of mental health and disorder. Specifically, we examined how people's health and disorder judgments were impacted by condition source, individual valuation, group valuation, and outcome. Our findings carry significant implications for understanding both the relational issue and the evaluative issue. Moreover, they shed light on how ordinary views align with philosophical perspectives (i.e., positivism, negativism, naturalism, normativism, and hybridism) and have implications for public health measures and clinical psychiatry.

We observed that people's health and disorder judgments are both associated with their dysfunction judgements: higher dysfunction judgments are associated with higher disorder judgments and lower health judgments. We also observed that people's health judgments are influenced by individual valuation and outcomes, whereas their disorder judgments are not. Instead, disorder judgments are influenced by condition source. Overall, the lay conception of mental health appears to be both positive and normativist, while the lay conception of mental disorder aligns with a naturalist perspective, at least to the extent that dysfunction plays an important role in categorizing a condition as a disorder. However, our finding that people's dysfunction judg-

ments are influenced by individual valuation and outcomes poses a strong challenge to naturalist accounts.

## Appendix

This supplementary section serves to complement our main discussion of people's concepts of mental health and disorder and the factors that influence people's judgments about whether a condition constitutes a mental disorder or a healthy state. It contains exploratory findings that, while not essential to the primary conclusions of the paper, contribute to contextualizing our main findings and to offering a more comprehensive picture. They also highlight possible future research directions.

The effect of source, outcome, individual valuation, and group valuation on participant's responsibility judgments we ran an ANOVA. Gender, age, ethnicity, political ideology, and possibility were entered as covariates. These same covariates were entered into each of the analyses which follow. The result of this analysis revealed a significant main effect of source,  $F(1, 326)=15.591, p<.001, \eta_p^2=0.046$ , outcome,  $F(1, 326)=18.480, p<.001, \eta_p^2=0.054$ , individual valuation,  $F(1, 326)=23.292, p<.001, \eta_p^2=0.067$ , and the covariates possibility,  $F(1, 326)=16.627, p<.001, \eta_p^2=0.049$ , and political ideology,  $F(1, 326)=13.185, p<.001, \eta_p^2=0.039$ .

The main effect of source was that participant's responsibility judgments were significantly lower in the genetic cases ( $M=4.65, SD=1.55$ ) than in the upbringing cases ( $M=5.31, SD=1.56$ ). The main effect of outcome was that participant's responsibility judgments were significantly lower in the negative cases ( $M=4.62, SD=1.56$ ) than in the positive cases ( $M=5.34, SD=1.56$ ). The main effect of individual valuation was that participant's responsibility judgments were significantly lower in the negative cases ( $M=4.578, SD=1.54$ ) than in the positive cases ( $M=5.38, SD=1.55$ ). The effect of possibility was that higher possibility judgments were associated with higher responsibility judgments. The effect of political ideology was that higher conservatism was associated with higher responsibility judgments.

Next, for participant's goal judgments an ANOVA revealed a significant main effect of outcome,  $F(1, 326)=392.669, p<.001, \eta_p^2=0.546$ , and individual valuation,  $F(1, 326)=15.327, p<.001, \eta_p^2=0.045$ .

The main effect of outcome was that participant's goal judgments were significantly lower in the positive cases ( $M=2.39, SD=1.55$ ) than in the negative cases ( $M=5.72, SD=1.55$ ). The main effect of individual valuation was that participant's responsibility judgments were significantly lower in the positive cases ( $M=3.73, SD=1.54$ ) than in the negative cases ( $M=4.38, SD=1.54$ ).

Then, for participant's control judgments an ANOVA revealed was a significant main effect of source,  $F(1, 326)=17.610, p<.001, \eta_p^2=0.051$ , outcome,  $F(1, 326)=11.656, p<.001, \eta_p^2=0.035$ , individual valuation,  $F(1, 326)=16.375, p<.001, \eta_p^2=0.048$ , and the covariate possibility,  $F(1, 326)=26.003, p<.001, \eta_p^2=0.074$ .

The main effect of source was that participant's control judgments were significantly lower in the genetic cases ( $M=3.12, SD=1.59$ ) than in the upbringing cases ( $M=3.85, SD=1.60$ ). The main effect of outcome was that participant's control judgments were significantly lower in the negative cases ( $M=3.19, SD=1.60$ ) than in

the positive cases ( $M=3.78$ ,  $SD=1.60$ ). The main effect of individual valuation was that participant's control judgments were significantly lower in the negative cases ( $M=3.14$ ,  $SD=1.60$ ) than in the positive cases ( $M=3.83$ ,  $SD=1.59$ ). The effect of possibility was that higher possibility judgments were associated with higher control judgments.

Next, for participant's author judgments an ANOVA revealed a significant main effect of source,  $F(1, 326)=8.845$ ,  $p=.003$ ,  $\eta_p^2=0.026$ , outcome,  $F(1, 326)=11.656$ ,  $p<.001$ ,  $\eta_p^2=0.035$ , individual valuation,  $F(1, 326)=13.419$ ,  $p<.001$ ,  $\eta_p^2=0.040$ , and the covariates, possibility,  $F(1, 326)=27.830$ ,  $p<.001$ ,  $\eta_p^2=0.079$ , and political ideology,  $F(1, 326)=12.607$ ,  $p<.001$ ,  $\eta_p^2=0.037$ . There was also a significant two-way interaction between outcome and group valuation,  $F(1, 326)=8.524$ ,  $p=.004$ ,  $\eta_p^2=0.025$ .

The main effect of source was that participant's author judgments were significantly lower in the genetic cases ( $M=3.89$ ,  $SD=1.66$ ) than in the upbringing cases ( $M=4.42$ ,  $SD=1.65$ ). The main effect of outcome was that participant's author judgments were significantly lower in the negative cases ( $M=3.85$ ,  $SD=1.67$ ) than in the positive cases ( $M=4.46$ ,  $SD=1.65$ ). The main effect of individual valuation was that participant's author judgments were significantly lower in the negative cases ( $M=3.83$ ,  $SD=1.65$ ) than in the positive cases ( $M=4.48$ ,  $SD=1.64$ ). The effect of possibility was that higher possibility judgments were associated with higher author judgments. The effect of political ideology was that higher conservatism was associated with higher author judgments.

Simple effects tests with Bonferroni correction were performed on the two-way interaction between outcome and group valuation. First, for positive outcome cases, there was no significant difference in participant's author judgments between the positive group valuation case ( $M=4.21$ ,  $SD=1.66$ ) and the negative group valuation case ( $M=4.72$ ,  $SD=1.66$ ). Nor was there any significant difference for negative outcome cases between participant's author judgments in the positive group valuation case ( $M=4.11$ ,  $SD=1.67$ ) than in the negative individual valuation case ( $M=3.59$ ,  $SD=1.66$ ). Second, for positive group valuation cases, there was no significant difference in participant's author judgments between the positive and the negative outcome cases. In contrast, for negative group valuation cases, participant's author judgments were significantly higher in the positive outcome case than in the negative outcome case ( $p<.001$ ).

Finally, for participant's well-being judgments an ANOVA revealed a significant main effect of outcome,  $F(1, 326)=132.072$ ,  $p<.001$ ,  $\eta_p^2=0.288$ , individual valuation,  $F(1, 326)=64.346$ ,  $p<.001$ ,  $\eta_p^2=0.165$ , group valuation,  $F(1, 326)=8.756$ ,  $p=.003$ ,  $\eta_p^2=0.026$ , and the covariate possibility,  $F(1, 326)=14.976$ ,  $p<.001$ ,  $\eta_p^2=0.044$ .

The main effect of outcome was that participant's well-being judgments were significantly lower in the negative cases ( $M=3.30$ ,  $SD=1.32$ ) than in the positive cases ( $M=4.95$ ,  $SD=1.32$ ). The main effect of individual valuation was that participant's well-being judgments were significantly lower in the negative cases ( $M=3.55$ ,  $SD=1.32$ ) than in the positive cases ( $M=4.69$ ,  $SD=1.32$ ). The main effect of group valuation was that participant's well-being judgments were significantly lower in the negative cases ( $M=3.91$ ,  $SD=1.33$ ) than in the positive cases ( $M=4.33$ ,  $SD=1.36$ ).

The effect of possibility was that higher possibility judgments were associated with higher well-being judgments.

**Funding** Open access funding provided by Aarhus Universitet.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bachmann, L. M., Mühleisen, A., Bock, A., ter Riet, G., Held, U., & Kessels, A. G. (2008). Vignette studies of medical choice and judgement to study caregivers' medical decision behaviour: Systematic review. *BMC Medical Research Methodology*, *8*, 1–8.
- Béghin, G., & Faucher, L. (2023). Does the Lay Concept of Mental Disorder Necessitate a dysfunction? In K. Hens, & de A. Block (Eds.), *Advances in experimental philosophy of Medicine* (pp. 71–96). Bloomsbury.
- Bishop, F., & Yardley, L. (2010). The development and initial validation of a new measure of lay definitions of health: The wellness beliefs scale. *Psychology and Health*, *25*(3), 271–287.
- Blaxter, M. (1990). *Health and lifestyles*. Routledge.
- Blaxter, M. (2010). *Health*. Polity.
- Boorse, C. (1977). Health as a theoretical concept. *Philosophy of Science*, *44*, 542–573.
- Boorse, C. (1997). A rebuttal on health. In J. M. Humber, & R. F. Almeder (Eds.), *What is disease?* (pp. 3–143). Humana.
- Boorse, C. (2014). A second rebuttal on health. *Journal of Medicine and Philosophy*, *39*(6), 683–724.
- Calnan, M. (1987). *Health and illness: The lay perspective*. Tavistock Publications.
- Clements, F. E. (1932). *Primitive concepts of disease*. University of California Press.
- Cooper, R. (2002). Disease. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *33*(2), 263–282.
- Cooper, R. (2005). *Classifying madness*. Springer.
- De Block, A., & Hens, K. (2021). A plea for an experimental philosophy of medicine. *Theoretical Medicine and Bioethics*, *42*(3–4), 81–89.
- Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., & Reed, G. M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical and Health Psychology*, *15*(2), 160–170.
- Foster, G. M. (1976). Disease etiologies in non-western medical systems. *American Anthropologist*, *78*(4), 773–82.
- Fulford, K. W. M. (1989). *Moral theory and medical practice*. Cambridge University Press.
- Fuller, J. (2018). What are chronic diseases? *Synthese*, *196*(7), 3197–3220.
- Giroux, É. (2016). Introduction. Why a book on naturalism in the philosophy of health? *Naturalism in the philosophy of health: Issues and implications* (pp. 1–16).
- Glackin, S. N. (2019). Grounded disease: Constructing the social from the biological in medicine. *The Philosophical Quarterly*, *69*(275), 258–276.
- Graham, G. (2010). *The disordered mind: An introduction to philosophy of mind and mental illness*. Routledge.
- Griffiths, P. E., & Stotz, K. (2008). Experimental philosophy of science. *Philosophy Compass*, *3*(3), 507–521.

- Grol-Prokopczyk, H., Freese, J., & Hauser, R. M. (2011). Using anchoring vignettes to assess group differences in general self-rated health. *Journal of Health and Social Behavior*, 52(2), 246–261.
- Herzlich, C. (1973). *Health and illness*. Academic.
- Hughner, R. S., & Kleine, S. S. (2004). Views of health in the lay sector: A compilation and review of how individuals think about health. *Health*, 8(4), 395–422.
- Jensen, L. A., & Allen, M. N. (1994). A synthesis of qualitative research on wellness-illness. *Qualitative Health Research*, 4(4), 349–369.
- Kingma, E. (2019). Contemporary accounts of health. In *Health* (pp. 289–318).
- Kirk, S. A., Wakefield, J. C., Hsieh, D., & Pottick, K. (1999). Social context and social workers' judgment of mental disorder. *Social Service Review*, 73(1), 82–104.
- Latham, A. J., & Varga, S. (forthcoming). Is “dysfunction” a value-neutral concept?
- Lemoine, M. (2013). Defining disease beyond conceptual analysis: An analysis of conceptual analysis in philosophy of medicine. *Theoretical Medicine and Bioethics*, 34, 309–325.
- Machery, E. (2023). The folk concept of disease. In K. Hens, & de A. Block (Eds.), *Advances in experimental philosophy of medicine* (pp. 51–70). Bloomsbury.
- Matthewson, J., & Griffiths, P. E. (2017). Biological criteria of disease: Four ways of going wrong. *Journal of Medicine and Philosophy*, 42(4), 447–466.
- McKague, M., & Verhoef, M. (2003). Understandings of health and its determinants among clients and providers at an urban community health center. *Qualitative Health Research*, 13(5), 703–717.
- Murphy, D. (2021). Concepts of disease and health. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2021 Edition). <https://plato.stanford.edu/archives/spr2021/entries/health-disease/>
- Nordenfelt, L. (1995). *On the nature of health: An action-theoretic approach*. Springer Science & Business Media.
- Nordenfelt, L. (2007). The concepts of health and illness revisited. *Medicine Health Care and Philosophy*, 10, 5–10.
- Nordenfelt, L. (2017). On concepts of positive health. In *Handbook of the philosophy of medicine* (pp. 29–43).
- Payton, K. S., & Gould, J. B. (2023). Vignette research methodology: An essential tool for quality improvement collaboratives. *Healthcare* (Vol. 11). Multidisciplinary Digital Publishing Institute (MDPI).
- Radden, J. (2019). Mental disorder (illness). In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2019 Edition). <https://plato.stanford.edu/archives/win2019/entries/mental-disorder/>
- Reiss, J., & Ankeny, R. A. (2016). Philosophy of medicine. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition). <https://plato.stanford.edu/archives/spr2022/entries/medicine/>
- Rogers, S. L. (1944). Disease concepts in North America. *American Anthropologist*, 46(4), 559–64.
- Schwartz, P. H. (2007). Decision and discovery in defining ‘disease’. In *Establishing medical reality: Essays in the metaphysics and epistemology of biomedical science* (pp. 47–63).
- Schwartz, P. H. (2017). Progress in defining disease: Improved approaches and increased impact. *Journal of Medicine and Philosophy*, 42, 485–502.
- Sholl, J. (2015). Escaping the conceptual analysis straitjacket: Pathological mechanisms and Canguilhem's biological philosophy. *Perspectives in Biology and Medicine*, 58, 395–418.
- Troisi, A., & Dieguez, G. (2022). Etiology of psychiatric disorders: Lay beliefs and the role of gender, field of study and personality traits. *Clinical Neuropsychiatry*, 19(3), 150–155.
- Tse, J. S., & Haslam, N. (2023). What is a mental disorder? Evaluating the lay concept of mental illhealth in the United States. *BMC Psychiatry*, 23(1), 224.
- Varga, S., Latham, A. J., & Machery, E. (forthcoming). *Health and disease: Experimental philosophy of medicine*.
- Varga, S., & Latham, A. J. (forthcoming). *Is health the absence of disease?*
- Venkatapuram, S. (2011). *Health justice: An argument from the capabilities approach*. Polity Press.
- Venkatapuram, S. (2013). Health, vital goals, and central human capabilities. *Bioethics*, 27(5), 271–279.
- Wakefield, J. C. (1992). The concept of mental disorder. *American Psychologist*, 47, 373–388.
- Wakefield, J. C. (2005). On winking at the facts, and losing one's hare: Value pluralism and the harmful dysfunction analysis. *World Psychiatry*, 4(2), 88–89.
- Wakefield, J. C. (2007). The concept of mental disorder: Diagnostic implications of the harmful dysfunction analysis. *World Psychiatry*, 6(3), 149.
- 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(6), 648–682.

- Wakefield, J. (2021). Do the empirical facts support the harmful dysfunction analysis? Reply to Luc Faucher. In L. Faucher, & D. Forest (Eds.), *Defining mental disorder: Jerome Wakefield and his critics* (pp. 71–96). MIT Press.
- Wakefield, J. C., & Conrad, J. A. (2020). Harm as a necessary component of the concept of medical disorder: Reply to Muckler and Taylor. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, 45(3), 350–370.
- Wakefield, J. C., Kirk, S. A., Pottick, K. J., Tian, X., & Hsieh, D. K. (2006). The lay concept of conduct disorder: Do non-professional use syndromal symptoms or internal dysfunction to distinguish disorder from delinquency? *Canadian Journal of Psychiatry*, 51(4), 210–217.
- Weller, S. C. (1984). Cross-cultural concepts of illness: Variation and validation. *American Anthropologist*, 86(2), 341–351.
- Wilkinson, S. (2023). *Philosophy of psychiatry: A contemporary introduction*. Routledge.
- Wren-Lewis, S., & Alexandrova, A. (2021). Mental health without well-being. *The Journal of Medicine and Philosophy*, 46(6), 684–703.
- Young, J. (1978). Illness categories and action strategies in a Tarascan Town. *American Ethnologist*, 5, 81–97.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.