EUROPEAN JOURNAL OF ANALYTIC PHILOSOPHY

UDC 101 | ISSN 1845-8475

Vol. 17, No. 2, 2021

REGULAR ARTICLES

FAMINE, AFFLUENCE, AND AMORALITY, **David Sackris** | LOGICAL RELATIVISM THROUGH LOGICAL CONTEXTS, **Jonas R. Becker Arenhart**

BOOK SYMPOSIUM ON THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE

NTRODUCTION BY GUEST EDITORS, **Maria Cristina Amoretti and Elisabetta Lalumera** | FROM ENGEL TO ENACTIVISM: CONTEXTUALIZING THE BIOPSYCHOSOCIAL MODEL, **Awais Aftab and Kristopher Nielsen** | CENTRIFUGAL AND CENTRIPETAL THINKING ABOUT THE BIOPSYCHOSOCIAL MODEL IN PSYCHIATRY, **Kathryn Tabb** | HOW TO BE A HOLIST WHO REJECTS THE BIOPSYCHOSOCIAL MODEL, **Diane O'Leary** | CAUSATION AND CAUSAL SELECTION IN THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE, **Hane Htut Maung** | THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE: RESPONSES TO THE 4 COMMENTARIES, **Derek Bolton**

SPECIAL ISSUE: PHILOSOPHY OF MEDICINE

INTRODUCTION BY GUEST EDITORS, **Saana Jukola and Anke Bueter** | DIAGNOSTIC JUSTICE: TESTING FOR COVID-19, **Ashley Graham Kennedy and Bryan Cwik** | ADAPT TO TRANSLATE – ADAPTIVE CLINICAL TRIALS AND BIOMEDICAL INNOVATION, **Daria Jadreškić** | WRONGFUL MEDICALIZATION AND EPISTEMIC INJUSTICE IN PSYCHIATRY: THE CASE OF PREMENSTRUAL DYSPHORIC DISORDER, **Anne-Marie Gagné-Julien** | MEDICALIZATION OF SEXUAL DESIRE, **Jacob Stegenga** | WHEN A HYBRID ACCOUNT OF DISORDER IS NOT ENOUGH: THE CASE OF GENDER DYSPHORIA, **Kathleen Murphy-Hollies** | THE QUANTITATIVE PROBLEM FOR THEORIES OF DYSFUNCTION AND DISEASE, **Thomas Schramme**

EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

UDC 101
ISSN (Print) 1845-8475
ISSN (Online) 1849-0514
https://doi.org/10.31820/ejap

Publisher, Editorial office
University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy
Address: Sveučilišna avenija 4, 51000 Rijeka, Croatia
Phone: +385 51 669 794
E-mail: eujap@ffri.uniri.hr
Web address: https://www.eujap.uniri.hr
Printed by Impress, Opatija (Croatia)

Publication of the journal is supported financially by
the Ministry of Science and Education of the Republic of Croatia

TABLE OF CONTENTS

REGULAR ARTICLES

FAMINE, AFFLUENCE, AND AMORALITY
David Sackris……..…….(A1)5

LOGICAL RELATIVISM THROUGH LOGICAL CONTEXTS
Jonas R. Becker Arenhart…….…….(A2)5

BOOK SYMPOSIUM ON THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE

INTRODUCTION BY GUEST EDITORS
Maria Cristina Amoretti and Elisabetta Lalumera…….(M1)5

FROM ENGEL TO ENACTIVISM: CONTEXTUALIZING THE BIOPSYCHOSOCIAL MODEL
Awais Aftab and Kristopher Nielsen…….(M2)5

CENTRIFUGAL AND CENTRIPETAL THINKING ABOUT THE BIOPSYCHOSOCIAL MODEL IN PSYCHIATRY
Kathryn Tabb…..……..……..…..……..……..…(M3)5

HOW TO BE A HOLIST WHO REJECTS THE BIOPSYCHOSOCIAL MODEL
Diane O'Leary..……..……(M4)5

CAUSATION AND CAUSAL SELECTION IN THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE
Hane Htut Maung……..(M5)5

THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE: RESPONSES TO THE 4 COMMENTARIES
Derek Bolton……..(M6)5

SPECIAL ISSUE: PHILOSOPHY OF MEDICINE

INTRODUCTION BY GUEST EDITORS
Saana Jukola and Anke Bueter…..……..(SI1)5

DIAGNOSTIC JUSTICE: TESTING FOR COVID-19
Ashley Graham Kennedy and Bryan Cwik……(SI2)5

ADAPT TO TRANSLATE – ADAPTIVE CLINICAL TRIALS AND BIOMEDICAL INNOVATION
Daria Jadreškić……..……..(SI3)5

WRONGFUL MEDICALIZATION AND EPISTEMIC INJUSTICE IN PSYCHIATRY: THE CASE OF PREMENSTRUAL DYSPHORIC DISORDER
Anne-Marie Gagné-Julien……..(SI4)5

MEDICALIZATION OF SEXUAL DESIRE
Jacob Stegenga……..……..……(SI5)5

WHEN A HYBRID ACCOUNT OF DISORDER IS NOT ENOUGH: THE CASE OF GENDER DYSPHORIA
Kathleen Murphy-Hollies..…….……(SI6)5

THE QUANTITATIVE PROBLEM FOR THEORIES OF DYSFUNCTION AND DISEASE
Thomas Schramme……..……..(SI7)5

ABSTRACTS (SAŽECI).…..……..(AB)5

AUTHOR GUIDELINES AND MALPRACTICE STATEMENT……..…….(AG)5

EuJAP | Vol. 17 | No. 2 | 2021
UDC: 17.022:17.034
https://doi.org/10.31820/ejap.17.2.1

FAMINE, AFFLUENCE, AND AMORALITY

David Sackris[1]

[1]Arapahoe Community College

Original scientific article – Received: 14/07/2021 Accepted: 18/09/2021

ABSTRACT

I argue that the debate concerning the nature of first-person moral judgment, namely, whether such moral judgments are inherently motivating (internalism) or whether moral judgments can be made in the absence of motivation (externalism), may be founded on a faulty assumption: that moral judgments form a distinct kind that must have some shared, essential features in regards to motivation to act. I argue that there is little reason to suppose that first-person moral judgments form a homogenous class in this respect by considering an ordinary case: student readers of Peter Singer's "Famine, Affluence, and Morality". Neither internalists nor externalists can provide a satisfying account as to why our students fail to act in this particular case, but are motivated to act by their moral judgments in most cases. I argue that the inability to provide a satisfying account is rooted in this shared assumption about the nature of moral judgments. Once we consider rejecting the notion that first-person moral decision-making forms a distinct kind in the way it is typically assumed, the internalist/externalist debate may be rendered moot.

Keywords: *Meta-ethics; moral judgment; internalism; externalism; natural kinds*

Correspondence: david.sackris@gmail.com

Introduction

Most academic philosophers have taught a class on Peter Singer's 1972 article 'Famine, Affluence, and Morality' at least once. In his essay, Singer critically assess the lifestyle of modern Westerners, illustrating how easily we could save the lives of the desperately poor if we would only choose to forgo trivial enjoyments, for example, exchanging our daily $5 latte for a 25¢ cup of Folger's, while donating the remainder to charity. Surely the life of a human being is more important than the momentary pleasure of a latte. Therefore, Singer posits, one is morally required to donate that remaining $4.75 to famine relief and make do with the less enjoyable good.

Singer's central argument is exceedingly simple and, prima facie, difficult to rebut (especially for introductory level students).[1] Typically, a substantial group of students will say that they think Singer is right, concluding that Westerners should do more to alleviate global suffering. But here is the rub: very few students seem to actually change their lifestyle one iota as a result of Singer's argument.[2]

Especially illustrative of this phenomenon is the class discussion of the central thought experiment in Singer's article. It goes like this: suppose you are walking down the street and see a small child drowning in a shallow pond. Surely you would feel morally obligated to save the child, even if it meant ruining the pants you were wearing. The value of the pants pales in comparison to the life of a human being who needs help through no fault of their own (1972, 231). The overwhelming majority of the students tend to agree with Singer that it would be *morally wrong* not to help the child, and a significant number even suggest that they would be willing to jail any person who ignores the drowning child and walks by. However, when the conversation moves to the starving children of East Bengal, students typically become less sure about the wrongness of not helping. Roughly, most students think that it would be good to help such children, and that people ought to do so, yet students rarely express the opinion that not helping is a significant moral wrong or that non-helpers belong in jail. They fail to express this opinion even though these same students are typically unable to poke significant holes in Singer's reasoning that the starving children of East Bengal are not relevantly different from a child drowning right in front of them. After lengthy discussion, some students reject Singer's ultimate conclusion that they are morally obligated

[1] And perhaps it can't be rebutted because it's a sound argument. It is not my aim to discuss the merits of Singer's argument here, but instead use it as an illustrative example.

[2] Admittedly, a small number of students are convinced by Singer's argument and do act on their newfound judgment; the exceptions are so notable that Nicholas Kristoff (2015) wrote a column about it.

to help the children of East Bengal without any real reason; many more appear to accept his conclusion but do nothing to conform their behavior to their newly formed judgment.[3]

This phenomenon of being intellectually *convinced* by a moral philosophical argument, yet seemingly *unmotivated* to behave according to one's conviction, appears to count as another piece of evidence in the long-standing philosophical dispute over the nature of moral judgment and motivation known as the *externalist vs. internalist* debate. *Externalists* hold that there is a basic disconnect between beliefs and behavioral motivation. Moral judgments, externalists claim, are not in themselves motivating. And we might agree that when discussing Singer's article, our students' beliefs and behaviors (or lack thereof) lend strong empirical support for such a position. The problem with simply accepting externalism, however, is that it is also clearly true that many moral judgments are, as a matter of fact, motivating: people typically act on their considered moral judgments.[4] In fact, this is precisely what *internalists* have traditionally maintained: one cannot make a *real* moral judgment without being motivated to act.[5] In this respect, internalism serves as a kind of 'best explanation' of typical human behavior.

Most likely our students would act on their moral judgment that they ought to save the drowning child right in front of them; I also can't deny that most students fail to act, and do not appear to be strongly motivated to act, on their in-class judgments about famine relief. The question then is this: how do we make sense of such mixed evidence, not from a normative standpoint but from a *descriptive* one?[6] That is, how do we account for the clearly observable phenomena of ordinary moral judgments whereby some moral judgments are highly motivating, almost always resulting in action, and other moral judgments do not result in any action or *even any apparent motivation to act*?

I argue that if we aim to account for real-world ethical decision-making by ordinary people, we should reconsider the internalist/externalist debate and entertain the possibility that neither view, *by itself*, is able to offer the correct account. Through an explicit consideration of this curious case, I aim to raise the following, neglected possibility: What if moral judgments do not form a distinct *kind*, at least in respect to motivational impact? I

[3] King (2018, 635) also makes the latter observation about her students and their reading of 'Famine, Affluence, and Morality'.
[4] Barring some other, overriding obligation.
[5] Both externalism and internalism will be carefully considered and defined in subsequent sections.
[6] It seems that theories of moral judgment are often about how moral judgments ought to or should be made (i.e., they are prescriptive), but the point here is that we should focus more on the observable behavior of ordinary decision-makers.

ultimately conclude that we have good reason to reconsider the view that *all* moral judgments will be either necessarily motivating or motivationally inert. That is, there may be different kinds of judgments that we classify as 'moral' yet, despite this ordinary language classification, it is not the case that these judgments will have all the same significant properties.

I begin by considering what the internalist and externalist might say about our student readers of 'Famine, Affluence, and Morality' and why their likely analyses of the situation are unsatisfactory. I then turn to what appears to be a shared, unargued for assumption of both internalists and externalists: that moral judgments form a distinctive kind and have necessary, shared features. I then argue that such an assumption should be reconsidered at least in respect to motivational features.[7] Reconsidering this assumption could lead to a resolution of the externalist/internalist debate.

1. What the internalist has to say about our students' judgments and behaviors

Let's suppose, for a moment, that moral judgments are necessarily motivating (i.e., that some version of internalism is true).[8] Obviously, the majority of students are not acting on their considered moral judgments in this case. Further, they do not *appear* to be highly motivated to act on said judgments; there are almost no barriers to their acting—they could donate through their smart phones immediately after class—yet they still typically fail to act.[9] One of the central difficulties with this debate is that it is nearly impossible to determine whether someone is at least minimally motivated by their judgment even when they fail to act on it. Given these facts, what must the internalist say about our students? We have three options:

a) Most students are practically irrational.
b) Most students are not making "real" moral judgments.
c) Most students experience some minimal motivation that does not arise to the level of action.[10]

[7] Few contemporary authors have questioned this assumption that moral judgments form a distinct kind. Sinnott-Armstrong and Thalia (2012, 2014), and Stich (2006) constitute exceptions.
[8] Internalism is both interpreted as a conceptual truth and as an empirical one. For example, Smith (1994) is essentially defending a defeasible conceptual connection, and Brink (1986) argues that if an amoralist is merely conceptually possible, then internalism is defeated. Prinz (2007) and Björnnson (2002) offer empirical arguments for internalism. An exact definition of internalism is difficult to pin down; for an overview, see Smith (1994, chapter 3) and Korsgaard (1986).
[9] *Almost* is the key word here. I assume that most American college students can spare a few dollars for famine relief at least once in a while.
[10] King (2018, 636) also lists these as the three likely responses for the internalist.

We begin with (a). For the students to be considered practically irrational, it must be the case that they are *not at all* motivated by their moral judgment. The *rational internalist* (henceforth, rationalism), for example, maintains that the recognition of a moral requirement provides a reason for action, and that such reasons motivate. Acting, or being motivated to act, on the recognition of such normative reasons is a requirement of rationality, and so rational individuals will be motivated to act on their moral judgments, barring instances of practical irrationality (Smith 1994). On a position like Smith's, barring the possibility that the students have some other, overriding moral obligation that conflicts with contributing to famine relief, we are led to conclude that the vast majority of our students are practically irrational. Let's see why.

Here is how Smith describes his internalist position:

> If an agent believes that she has a normative reason to φ, then she should rationally desire to φ. (Smith 1994, 148)

Smith accepts that there is a defeasible connection between our judgments and actions; namely, we don't *always* act on our moral judgments. This requires an explanation. He states:

> If an agent judges that it is [morally] right for her to φ in circumstances C, then either she is motivated to φ in C or she is practically irrational. (Smith 1994, 61).

By 'practically irrational' Smith means individuals who 'judge it right to act in various ways' but fail to act on those judgments (Smith 1994, 61). Such individuals must be suffering from 'weakness of will and other similar forms of practical unreason on their motivations' (Smith 1994, 61). If an individual is not motivated by what she considers a reason for action, then 'she fails to be rational by her own lights' (Smith 1994, 62). So, if a student judges that Singer has made a convincing argument, yet fails to be motivated to act on this judgment, then they are practically irrational.

To write off the majority of our students as practically irrational seems a bit too quick: we shouldn't rush to embrace a norm of rationality that does not fit the majority of seemingly rational individuals' reasoning and subsequent behavior.[11] Prima facie, my experience teaching ethics seems like an objection to Smith's argument: here are seemingly rational individuals who understand Singer's reasons (and have good reason to try

[11] Williams thinks it is too quick as well, his point being that by the students' own lights they are acting rationally (1979, 25). Smith (1994) aims to refute this claim. See especially chapter 5.

and understand his reasons, given that they will be tested on the material), accept them, yet seemingly fail to be motivated to act. But these same individuals are motivated to act on their moral judgments in many other routine situations, e.g., tracking down a fellow student who left their textbook in the classroom. Performing such an act may require more work than donating to charity, which can be accomplished via one's smartphone.[12]

Part of the problem with Smith's position, and many accounts of normative judgment like his, is that it is sometimes unclear what the project is supposed to be: a descriptive one or a prescriptive one.[13] Sadler (2003) astutely points this out. Is Smith's theory an analysis of the concept 'moral judgment' as used by an ideal agent, i.e., is it a theory about the nature of judgments as made by good and strong-willed persons, or is it meant to be an analysis of the concept as employed by ordinary individuals? It seems clear that he aims to do the latter.[14] Yet his account fails to explain what is going on in the typical ethics course, unless he wants to call the majority of undergraduates, and, I would contend, the majority of human beings, practically irrational. There would be no internalist/externalist debate if it didn't seem possible, in a very ordinary kind of way, to make a moral judgment without necessarily feeling motivated to act on said judgment. So, it is hard to see how failing to be motivated deserves the charge of practical irrationality.[15]

So, on a rationalist account like Smith's, in order to explain why most students fail to be strongly motivated to act on their judgment that more should be done for the starving children of East Bengal, we have to either accept that the majority of people are practically irrational even in contexts of careful deliberation, like a philosophy classroom, or accept that (a) does not appear to offer a satisfying analysis of our student's failure to act. The latter seems like the more plausible conclusion.

Let's now consider (b): our students are not making 'real' moral judgments. Instead of maintaining that our students are practically irrational or suffer from a contagious case of weakness of will, the

[12] Even if the reader is unsure of what to make of our student readers' judgments and for that reason dislikes my focus on this example, the phenomena of intellectually judging an act to be morally obligatory yet failing to actually carry it out does not seem to be all that unusual. The judgments we make concerning what we ought to do while lying awake at night are often not the ones we follow through on in the morning.

[13] Similarly, it is unclear whether Carroll's (2015) theory of aesthetic experience is meant to be a descriptive or prescriptive one. See Sackris and Larsen (2020).

[14] See especially Smith (1994, chapters 1 and 2).

[15] Setiya (2004) points out that even it if it is true that the concept 'moral judgment' necessarily includes motivation, if coming to see this requires significant philosophical reflection, then it is hardly fair to call those who fail to realize this 'irrational'.

internalist could maintain that the students are not making 'real' moral judgments. If motivation is part of the concept 'moral judgment', then lack of motivation might indicate that the concept is not actually being deployed. Rosati (2016) emphasizes the connection between failure to act and insincerity: '[I]f an individual makes a moral judgment, she is, ceteris paribus, motivated; if she is not motivated, she was not making a sincere and competent moral judgment at all, appearances to the contrary notwithstanding'.[16] So, if the students aren't motivated to act, then we might conclude that they are merely saying what they think we, their professors, want to hear, or that they have some other reason for falsely reporting their agreement with Singer.

There are additional considerations. It may be true that they have limited ability to act *in class at the moment* of the discussion of famine relief, so in that sense the critical reader may think this is a poor example. However, I ask students if they plan to go out and do anything differently (planning to act differently would seem to indicate current motivation), and the next class bring up the same sorts of questions: has anyone forgone their daily Starbuck's latte in favor of famine relief? Has anyone, instead of paying their fraternity dues, considered donating those dues to famine relief? Perhaps the chorus of 'Nos' supports the contention that they haven't made real moral judgments.

Yet it is not clear why we should think that our students are not making 'real' moral judgments in this particular case, when we would be unlikely to say the same thing about other topics where it would be difficult for students to act *in any fashion* even if they wished to, e.g., we might ask our students whether they think the use of torture by the state is permissible. Here the internalist would likely complain that there is a significant difference between this case and my preferred example: unlike the issue of famine relief, it is virtually impossible for students to act on their judgments about state-sanctioned torture in or out of class;[17] nonetheless, they could still be motivated by such judgments. Their motivation is merely frustrated in the torture case. The problem is that we don't have any direct evidence that they are motivated and frustrated; such direct evidence is unavailable. To say that they *must* be motivated and that their motivations are merely frustrated when considering torture sounds a bit like assuming the very thing that is supposed to be proven—whether they are *actually* motivated by their in-class moral judgments.

[16] Harman offers a similar formulation (1977, 33), as does Blackburn (1984, 188).
[17] For the most part. Of course, they could organize protests, run for office, etc., but there is no single action they could easily take to bring about their judgment regarding state sanctioned torture.

Given this problem of opacity, most philosophers likely just assume, whether they are committed internalists or externalists, that students are making real moral judgments in our ethics classes, whatever the topic—whether or not they have the ability to act on their judgments. If we do not believe that our students are capable of genuine moral reflection and judgment in our classes, we should probably stop teaching ethics. So (b) probably isn't the right answer.

That leaves us with (c). Let's now consider whether the internalist should be attracted to a position on which all moral judgments are accompanied by some minimal motivation, but that motivation need not rise to a level at which the individual would be motivated *enough* to act, even in situations where there are no practical obstacles to acting. On this position, although the students who agree with Singer don't *do anything*, they are nonetheless minimally motivated by their judgments.

First, let's consider whether a rationalist should be attracted to such a position. To review, on Smith's position, if students have judged that Singer is right, then they should thereby be motivated to act. Smith has little to say about *degrees* of motivation: however, he routinely appeals to depression as an example of a practical irrationality that *completely extinguishes* one's motivation to act:

> It is a commonplace, a fact of ordinary moral experience, that practical irrationalities of various kinds—various sorts of 'depression' as [Michael] Stocker calls them [1979, 744]—can leave someone's evaluative outlook intact while removing their motivation altogether. (Smith 1994, 120-121)[18]

Appealing to a completely will-draining depression fails to get at the core suggestion in (c): that our students have some minimal motivation that accompanies their judgment, but that the motivation is simply not strong enough to get them to act. In the context of teaching 'Famine, Affluence, and Morality', it is unlikely that most of our students are suffering from a kind of global, will-draining form of depression; if that were the case, they likely wouldn't have even made it to class.

If one wants to make sense of a claim like (c), identifying moral motivation with emotion may seem to be a natural move. If one advocates for a sentimentalist theory of morality and holds, like Jesse Prinz (2007), that

[18] For additional examples of Smith focusing on completely debilitating forms of mental illness, see pages 123 and 125 of his (1994).

moral judgments are constituted by emotions, then one has good reason for being attracted to (c).[19] As Prinz states:

> If moral judgments contain moral concepts, and moral judgments have an emotional composition, then moral judgments motivate action, because emotions are motivational states. [Sentimentalism] entails internalism (…). (Prinz 2007, 102)

On this view, every moral judgment does in fact contain some minimal motivation, and our students are likely feeling *some* emotions as they read of the plight of individuals caught up in tragic circumstances. On this position, even in cases where students fail to act on their judgments, we still cannot conclude that *they weren't motivated at all*: given the sentimentalist definition of a moral judgment, we should assume they feel some minimal motivation. Furthermore, it would be exceedingly difficult to prove that there isn't some kind of minimal motivation that corresponds to their judgment. Therefore, internalism, on this interpretation, is true by default.

Yet such a position is also problematic: lacking direct access to the subjective states of moral decision makers, it is impossible to show that moral judgment is, or is not, always accompanied by minimal motivation when the only readily available evidence is whether the individual ultimately acts. Elinor Mason dubs a view along the lines of (c) 'Weakest Internalism'. She says

> The only difference between weakest internalism and externalism is that weakest internalism says that when there is a moral judgement there is always some level of motivation, however slight and ineffective…. The chief point of weakest internalism seems to be to satisfy the basic internalist intuition that it is odd to judge that you ought to do something and yet not be motivated at all. But without an independent argument for internalism, that intuition is not a good enough justification for adding the internalist clause to the theory. (Mason 2008, 144)

What Mason means by 'an independent argument for internalism', I suppose, is something like this: an empirical argument in favor of the internalist thesis. So, if, e.g., sentimentalism is true and moral judgments

[19] Additional modern advocates of sentimentalism include Nichols (2004), Gill and Nichols (2008), and Slote (2010).

are in fact composed (in some fashion) of emotional states, we would then need empirical evidence that moral emotions, or all emotions, contain some minimal amount of motivation. Do we have any such evidence along these lines?

We would need an argument that shows either of the following: 1) that there is only some small subset of emotions involved in moral judgments, all of those emotions are in fact motivating, *and* that there is no other basis for moral judgments; or 2) that *all* emotions are motivating *and* that that there is no other basis for moral judgment. It would be very difficult to empirically demonstrate the former, [20] and Prinz, one of the chief contemporary supporters of sentimentalism, founds his position on the latter. Additionally, that all emotions are motivating appears to be taken as a truism by many within the psychology community.[21] Prinz says:

> In order to act, we must be motivated. Emotions and motivation are linked. Emotions exert motivating force. There is clinical evidence that, without emotions, people feel no inclination to act. (Prinz 2007, 17-18)

Prinz goes on to cite a Damasio and Van Hoesen (1983) article that discusses individuals with a condition called akinetic mutism. Damasio and Van Hoesen theorize that such individuals lie completely motionless because they have sustained damage to specific regions of the brain responsible for emotions. Without the ability to feel emotions, these individuals lack motivation to act in any fashion.

I do not deny that many emotions play a key role in motivation, but does akinetic mutism prove that *all* emotions motivate? It may be that without any emotional faculties a person will not have any inclination to act, but this, by itself, does not show that all emotions motivate. That is, it could be true that some subset of emotions is required to motivate action while it is also true that some other emotions don't play a direct motivational role.[22] If it is possible that there are non-motivational emotions, it is also possible that those emotions constitute some moral judgments.

[20] Haidt identifies six moral foundations, and he associates those foundations with 'characteristic emotions' but he does not identify moral judgments with specific emotions, nor suggest that other emotions cannot play any role in the six moral foundations he identifies. See his (2012), especially chapters 6 and 7. See also Cameron et al. (2015).

[21] See for example Stangor and Walinga (2014, 441-442).

[22] Blasi (2001) criticizes the view that emotions are necessarily motivational. Additionally, in their ontology of emotion, Hasting et al. (2011) state that many emotions have action tendencies, but they do not include motivation to act in their definition of emotion.

Prinz's chief inspiration, Hume, also thought that some emotions may not have a motivational function:

> For pride and humility are pure emotions in the soul, unattended with any desire, and not immediately exciting us to action. But love and hatred are not compleated within themselves, nor rest in that emotion, which they produce, but carry the mind to something farther. (Hume 1896, 368)

So, it may be that without the ability to feel love and anger we wouldn't *do* anything at all, yet that still doesn't tell us that pride and humility necessarily motivate, and it is not abundantly clear that pride and humility are *not* moral emotions. If not all emotions motivate, this leaves open the possibility that there could be moral judgments that are composed of non-motivating emotions.

Currently, it is simply not possible to prove that all emotions motivate, nor is it possible to concretely pinpoint some subset of emotions that make up *all* moral judgments, so the common idea expressed by Prinz that sentimentalism entails internalism could be false. No doubt we have felt our emotions motivate us to action; however, we can also think of emotional states that seem to play no role in motivating action; postulating some action for the latter emotions to supposedly motivate comes across as ad hoc. E.g., what actions do awe, satisfaction, astonishment, or pride motivate? What action does a feeling of the sublime motivate? What actions do moods motivate, such as general feelings of depression or anxiety? It is hard to see how all of these states could be necessarily action-directing.

In this section I have argued that internalism does not seem adequate for explaining the behavior of our students and their consideration of Singer's argument. Internalists could try to maintain either that a) the vast majority of our students are practically irrational; b) the vast majority of our students do not make real moral judgments; or c) the vast majority of our students are at least minimally motivated.[23] I argued that there is little reason to think that in this particular case (but not in other, similar situations) that our students do not make real moral judgments; I further argued that if our students are practically irrational, then basically all normal adults are practically irrational, and if that is the case, then the charge of irrationality seems to lose its normative force. Although (c) strikes me as the most

[23] They could also maintain some combination of these three is occurring in the classroom, which is slightly more plausible: some students are minimally motivated, some students are amoralists, and some students aren't making real moral judgments. As I discussed, however, we have independent reasons to be skeptical of each possibility.

plausible response, it is problematic in that there is no way to show that individuals did in fact have some minimal motivation, and I offered reasons for rejecting the commonsense sentimentalist position that maintains that all emotions play a motivational role. At this point, it doesn't seem that the internalist theory, considered as a universal account of moral judgment, can offer a satisfying analysis of our students' behavior. Let us now turn to examining what the externalist has to say about the behavior and judgments of our students. To do so, we need to first examine what exactly the externalist believes.

2. What the externalist has to say about our students' judgments and behaviors

Externalists deny that there is an essential connection between making moral judgments and being motivated to act. Shafer-Landau (2000, 271) characterizes the position as little more than the rejection of internalism. The main idea is that a moral judgment is one thing, the motivation to act on that judgment is another; there is no necessary connection between a moral judgment and the desire to act. However, the externalist position is, in reality, more complicated than this. The rejection of internalism is typically conceptually connected to some other position that is simultaneously maintained, e.g., that moral judgments are always a kind of belief, and beliefs do not motivate; or that moral judgments are always the recognition of a moral fact, and the recognition of a fact does not motivate. For example, Brink (1986, 26) attacks the internalist thesis as part of a defense of moral realism and observes that many philosophers have maintained that moral realism and internalism are generally incompatible. In this respect, the externalist is just as committed to the idea that moral judgments form a distinct kind as the internalist is. What they disagree on is which significant features a judgment must have to be included in the class 'moral judgment'.

The chief argument in favor of externalism is merely an attempt to refute internalism, rendering externalism true by default. To refute internalism, the externalist typically appeals to a character known as the 'amoralist'. An *amoralist* is a hypothetical person described as someone who knows about moral values and makes moral judgments, but remains wholly unmotivated by them.[24] Shafer-Landau makes clear how important the amoralist is for the defenders of externalism:

[24] The following authors discuss the amoralist: Bedke (2008), Brink (1986), Bromwich (2013), Buckwalter and Turri (2017), King (2018), Nichols (2002), Smith (1994), Sadler (2003), Shafer-

> [The externalist] need defend only the conceptual possibility of an agent who on a single occasion fails to be motivated by a moral judgment that he endorses…. Establishing the possibility is all we need to undermine [internalism]; one doesn't show [that] internalism [is] true just by showing (if one can) that there are in fact no amoralists. [Internalism] is vindicated if and only if there cannot be any such people. (Shafer-Landau 2000, 271)

Whether there could in fact be such a person as an amoralist is itself recognized as a contentious thesis in the literature (Shafer-Landau 2000; Mason 2008). The contentiousness regarding whether such a person could even exist makes clear that the amoralist trope is an intuition pump that essentially replicates the original controversy. For whether one thinks that there could be such a thing as an amoralist is contingent on one's intuitions about the nature of moral judgment.[25] If one thinks that real moral judgments necessarily motivate (internalism), then one is likely to think that either there couldn't really be such a person as an amoralist, or that such a person, if they exist, isn't *really* making moral judgments, at least not in the same way that psychologically normal people do.[26] If, on the other hand, one thinks that moral judgments are not necessarily motivating (externalism), then one likely thinks that amoralists are possible, and that they very well might exist, say, in the form of a moral cynic or psychopath. Whether amoralists really are possible isn't all that important here, in part because the figure of the amoralist does not seem to have advanced the debate on the nature of moral judgments in any significant way,[27] and in part because our classrooms are unlikely to be populated by vast tracts of amoralists. If our students did not have any feelings *at all* about the issue of world hunger, they certainly wouldn't squirm in their seats when the instructor points out the frivolous things they gladly use their spending money on without a second thought instead of contributing to famine relief.[28]

Landau (2000), Sinnott-Armstrong (2014), and Svavarsdottir (1999). Further, the following authors consider the possibility that there are actual amoralists, namely, psychopaths: Kennett (2006), Matthews (2014), Maibom (2018), Nichols (2002), Smith (1994), and Sinnott-Armstrong (2014).

[25] And this difference in intuition may be traceable to the fact that different people hold slightly different, largely overlapping concepts of 'moral judgment'. See Francén (2010).

[26] A common internalist response to the amoralist example is to deny that amoralists are in fact making moral judgments in the same way as ordinary people, or even using moral language in the same way. See for example Hare (1952) and Smith (1994).

[27] See Francén (2010) and Rosati (2016) for a similar assessment.

[28] As discussed above, whether those feelings are motivational is a separate question. I don't doubt they felt something; I doubt all feelings motivate.

Putting aside, for now, the hypothetical individual who can reach moral judgments and be *wholly* unmoved by them, how does the externalist explain ordinary moral decision-making and the reliable connection between moral judgment and behavior that we typically find? In explaining this reliable connection, Brink says the following:

> Though it makes the motivational force of moral considerations a matter of contingent psychological fact, externalism can base this motivation on 'deep' or widely shared psychological facts…. [A]s a matter of contingent psychological fact, the vast majority of people will have at least a desire to comply (even) with other regarding moral demands. Moral motivation, on such a view, can be widespread and predictable, even if it is neither necessary, nor universal, nor overriding. (Brink 1986, 31)

Shafer-Landau also believes the connection between judgment and action will involve emotions and desires:

> The importance of any such account [of moral motivation] is that it makes the existence of the relevant desires contingent. This is easily seen when it comes to socialization stories, which explain the desires that constitute conscientious motivation as arising from early moral education and upbringing. On this line, it is conceptually possible for moral judgments to fail to motivate because it is conceptually possible for individuals either to receive a quite poor early training, or to receive a fine one and later distance themselves from it in fundamental respects. (Shafer-Landau 2000, 287)

Someone new to philosophy, but a critical thinker nonetheless, might read these two passages and think: Wait, what's the difference between externalism and internalism? Aren't these two groups telling the same story as to why people act on their moral judgments? The answer to the second question is: Yes, they are telling the same basic story. Externalists believe that emotions and desires do in fact reliably motivate people to act on their judgments, just as internalists do. The difference is that externalists maintain that said reliable motivation is *contingent*. Both groups maintain that people reliably act on their moral judgments, and that in some cases it *appears* that individuals can make moral judgments without being motivated. Of course, for the internalist, this is merely an appearance that can be explained away.

If all the externalist demands is that the philosophical community admit the *conceptual possibility* of an individual who makes a moral judgment and doesn't feel at all motivated by it, then I grant that conceptual possibility. But such a concession doesn't tell us anything at all about what is going on psychologically with any actual person when they make a moral judgment. What we should be interested in is whether such a person could exist, as that would actually tell us something about the ontology of moral judgment, and not merely the concept, which may fail to pick out any distinctive psychological process at all.

The externalist, then, finds herself in a similar position to the internalist when describing the behavior of our students. The externalist, as suggested by Brink and Shafer-Landau above, should be open to maintaining something quite similar to the internalist: because the students do likely feel some emotional response to Singer's article, and because externalists do in fact accept that emotions (contingently) motivate, when it comes to actual human beings (i.e., when excluding amoralists), moral judgments are at least minimally motivating. That is, they should be open to accepting something like weakest internalism as a fairly accurate descriptive account of human moral judgment.[29]

As Mason (2008, 144) points out, there is little meaningful difference between weakest internalism and externalism. Although Mason's point was that the internalist might as well adopt externalism, the argument seems to cut both ways. Once we grant the possibility that all moral judgments have some minimal motivational force, externalism seems to lose its appeal. For it seems that the externalist has to rule out an intuitive account of the failure to act by definition: the students are motivated by their judgments but not to a sufficient degree to give rise to action.[30] For merely being weakly motivated is a more plausible explanation for failure to act on one's moral judgments than the explanations readily available to the externalist: e.g., maintaining that most students are themselves flawed in some way (i.e., they are amoralists), or that they had a flawed moral education, a possible explanation put forth by Shafer-Landau. For attributing failure of motivation to poor upbringing actually appears to cede some ground to the internalist: for she could then maintain that those

[29] On a position like *community internalism* (Drier 1990; Tresan 2006, 2009) it isn't even necessary that every individual within a given community feels motivated by their moral judgments, just so long as such judgments are made within the context of a community where individuals are reliably so motivated. On such a position, amoralists are in fact possible. On this view, moral motivation is contingently related to judgment at the individual level, just not at the community level. Here we might wonder how this view differs from the externalism as presented by Brink (1986), where he readily admits that most people will be reliably motivated by their judgments.

[30] Thanks to an anonymous reviewer for encouraging me to frame the problem for the externalist in this way.

individuals who *fail to be motivated at all* never learned to make 'real' moral judgments, and 'real' moral judgments are always minimally motivating. Hence, the internalist thesis is saved.

3. Shared assumptions of internalists and externalists

The internalist/externalist debate appears to rest on a key assumption that is often left unstated: that moral judgments form a distinct kind, or category, of judgment; if they didn't form a distinct kind, it wouldn't make sense to wonder if all moral judgments had some set of shared, significant features. [31] This assumption has the following entailment: if moral judgments form a distinct category or constitute a natural kind, then we should be able to identify some significant features that distinguish it from other kinds of judgments, such that if a judgment doesn't have said features it can't be a moral judgment.

Although several philosophers have in fact attempted to define what constitutes a moral judgment, a number of philosophers believe that a definition cannot be given. [32] If philosophers consciously admit to themselves that it is difficult to specify whether a judgment constitutes a moral one beyond some core, indisputable cases, they likely shouldn't simultaneously maintain that the concept 'moral judgment' has some necessary, specifiable features. If we consider the possibility that moral judgments form a heterogeneous class, then we can begin to entertain the possibility that moral judgments made in some contexts always motivate, and when made in other contexts they fail to motivate, without also

[31] Michael Gill (2009) has also observed that most meta-ethical theorizing simply begins with the assumption that moral judgments form a uniform or distinct kind that admit of a single conceptual analysis; he further wonders whether we can determine ordinary speaker's meta-ethical commitments based on their usage of moral language. He also wonders if the concept 'moral judgment' is employed differently by different speakers, or differently by the same speaker in different contexts. In this paper, I am more concerned with the referent of 'moral judgment', i.e., does it actually pick out a distinct type or process of judging. Nonetheless, I believe Gill gives strong arguments against there being a uniform *use* of the concept.

[32] Shafer-Landau (2015), for example, does not believe 'morality' can be defined, which would seem to imply that 'moral judgment' is similarly undefinable. See his 'Introduction'. Smith (1994) takes it that there is a kind of commonsense understanding of 'moral judgment' such that it can be defined on the basis of moral platitudes. See his chapter 1; this idea will be subsequently challenged. Richardson (2018), in his *Stanford Encyclopedia of Philosophy* entry on moral reasoning states '[W]e will need to have a capacious understanding of what counts as a moral question. For instance, since a prominent position about moral reasoning is that the relevant considerations are not codifiable, we would beg a central question if we here defined "morality" as involving codifiable principles or rules'. Svavarsdottir admits that 'it is of course notoriously difficult to say what distinguishes moral judgments from other evaluative or normative judgments' (1999, n. 6). Drier states 'we should just admit that it may be vague whether a given judgment is moral or not' (1996, 411, n. 419). I don't deny that there are widely accepted paradigm examples of moral judgments. Nonetheless, 'moral judgment' is clearly not a sharply defined concept, which philosophers seem to readily recognize.

concluding that only one 'real' moral judgment was ultimately made. Do we have any strong reasons *in favor* of thinking that moral judgments do in fact form a distinct kind?

The chief evidence relied upon by contemporary meta-ethicists is generally drawn from observations about language use, and the position that moral judgments do form a distinct kind is rarely substantively, or directly, argued for.[33] Gill aptly summarizes the recent state of the field:

> Twentieth century meta-ethicists typically presented some examples of ordinary discourse. But they didn't gather data in any kind of comprehensive and systematic way…. *For if the concept of morality is sharply unitary and robustly determinate*—if the relevant meta-ethical information is encoded in the DNA of every use of moral terms—then one handful of commonsense judgments, intuitions, and platitudes will instantiate the same meta-ethical commitments as any other. (Gill 2009, 217, my italics)

However, as recent empirical work has shown, the assumptions of trained philosophers concerning the use of a concept have not always aligned with the thinking of non-philosopher language users.[34] For example, meta-ethicists have typically inferred from language use that ordinary speakers are moral absolutists.[35] Studies indicate this isn't true (Goodwin and Darley 2008; Beebe and Sackris 2016). Based on their research, Beebe and Sackris state 'Thus, we can see that because the strength of our participants' inclinations toward objectivism varies according to the issue in question, the question of whether they are moral objectivists is not going to have a simple "Yes" or "No" answer' (2016, 917). Perhaps we should then consider the possibility that the question as to whether moral judgments all motivate or all fail to motivate won't have a simple 'Yes' or 'No' answer either.

An analysis on which moral judgments do not form a distinct kind might explain, in part, why we have competing intuitions about the nature of moral judgment, and why certain kinds of cases trigger certain kinds of intuitions. We should notice that in the arguments for externalism, the key figure of the amoralist is rarely presented as failing to act on their considered moral judgment while directly confronted with the person who

[33] Kumar (2015) is the exception here.
[34] For example, studies seem to show that ordinary speakers do not always take justification as necessary for knowledge ascriptions. See Sackris and Beebe (2014).
[35] Smith, for example, states that "it is a platitude that our moral judgements at least purport to be objective" (1994, 84).

will be injured/victimized by their failure to act. Instead, they are typically presented as being convinced by a moral argument concerning some far-off issue/individual and failing to act, much like the situation of our students in relation to the individuals of East Bengal.
Consider the following three examples:

> Virginia has put her social position at risk to help a politically persecuted stranger because she thinks it is the right thing to do. Later she meets Patrick, who could, without any apparent risk to himself, similarly help a politically persecuted stranger, but who has made no attempt to do so. Our morally committed heroine confronts Patrick, appealing first to his compassion for the victims. Patrick rather wearily tells her that he has no inclination to concern himself with the plight of strangers… Patrick readily declares that he agrees with her moral assessment, but nevertheless cannot be bothered to help. (Svavarsdottir 1999, 176)

> Imagine an introductory philosophy student who has become convinced of the truth of a crude sort of ethical relativism. She believes that the ultimate moral standard comprises the fundamental mores of the society in which an action is performed. Armed with this view of morality, she issues certain moral judgments that she takes to be correct. But she is alienated from her society. Or, more likely, though she finds much of the pre-vailing cultural code amenable, she rejects a strand. She is voicing what she takes to be the moral truth, yet is unmoved. (Schaffer 2000, 274)

> Alice was raised to believe that the divine command theory is correct. That is, as Alice herself might say, she was raised to believe that our moral obligations are determined by the commands of God…. On the principle of an eye for an eye, Alice believes that capital punishment is obligatory in cases of murder, and she believes she has an obligation to support capital punishment. But she is deeply compassionate, and she is quite out of sympathy with what she takes to be God's vengefulness. Because of her compassion she is not motivated in the least to support capital punishment. She is in fact active in opposing it, even though she believes she is morally forbidden to do so. (Copp 1995, 190-191)

In my review of the literature, I have yet to find an argument in favor of externalism where an amoralist fails to act in response to some moral

dilemma that *directly* confronts them. That is, there is no fictional example offered in support of externalism like this:

> Bob is walking to campus to teach moral philosophy. Bob is completely proficient with moral terms and makes moral judgments all the time: of course, he understands morality—he holds the chair in moral philosophy! On the way to campus, Bob sees his neighbor's child drowning in a shallow pond. Bob could easily save the child by wading in and effortlessly plucking him out of the water. Bob knows that saving the child is the right thing to do, and judges it to be the right thing to do, yet he has no inclination to save the child. Besides, he doesn't want to be late for his own lecture on moral motivation. So, despite his judgment, Bob keeps walking.

Unsurprisingly, no one argues for externalism in this way. Perhaps this is because it is simply *implausible* to almost every party to the debate to imagine someone judging that it is right to save the drowning child right in front of them yet failing to be motivated by said judgment. What the Schaffer, Svavarsdottir, and Copp cases have in common is that the amoralist is not failing to help a desperate person right in front of them: they are merely failing to, in the case of the first example, help some abstract individual, and in the second and third cases they are failing to act on highly abstract moral judgments. In abstract, non-pressing cases of moral judgment, it may seem plausible that an individual could make such a judgment without being motivated. However, when confronted by a suffering person directly, it doesn't seem at all plausible that there could be an individual who makes a moral judgment yet fails to act. This result is suggestive. It suggests that moral judgments may not form a uniform and definitive class, at least when it comes to motivation: each side appeals to quite different examples, and the different examples yield differing intuitions, perhaps because our judgment processes are highly context-dependent.

4. Conclusion

The term 'moral judgment' has a lengthy philosophical history; nonetheless, I have argued here that we should entertain the possibility that this term does not pick out a naturally occurring category; that is, we should consider the possibility that the internalist/externalist debate is founded on the mistaken assumption that moral judgments constitute a distinctive kind. In fact, given the variety of objects and events that have been brought into the moral domain by human beings (especially in recent

history), we should tend in just the opposite direction: if what constitutes the moral is so diverse, perhaps moral judgments themselves form a diverse group.

What, then, should we say about our students? Undoubtedly most make judgments and feel *something* when reading the Singer article, but likely different students feel different emotions and form different beliefs, and some of those emotions/beliefs may not be motivating, or so minimally motivating that the label "internalism" becomes meaningless. But if *some* set of judgments do in fact always motivate individuals to action (for whatever reason—perhaps because in some cases moral judgments are primarily composed of strong emotions), then the externalist label is meaningless as well; for, as discussed, the externalist thesis is typically tied to other claims about the essential nature of moral judgment.

As stated at the beginning of the article, I have little doubt that almost every student would spring into action to save the drowning child right in front of them.[36] There are likely a great number of factors, including emotional ones, that would cause them to act in such a situation, but it may well be difficult to connect their motivation with any one particular factor. In that case, we should consider the possibility that moral judgment does not admit of a single, unified analysis, as well as the possibility that our concepts do not match up neatly with underlying psychological processes in a one-to-one fashion.

One final implication of the position argued for here to consider is that, if correct, we may no longer be able to draw a clear distinction drawn between 'bona fide' moral judgments and 'defective' ones; at the very least, moral judgments cannot be called defective on the grounds that they fail to motivate.[37] As we have seen, internalists are fond of drawing such a distinction to defend their position: for example, Prinz argues that psychopaths are unable to form real moral judgments because they are unable to feel emotions in the same way as typical individuals; for Prinz (2007, 42-47), their inability to be motivated by their moral judgements supplies evidence that moral judgments, as formed my average individuals, are in fact rooted in emotions and thereby reliably motivating.[38]

[36] I am thinking almost all would *at least* dial 911.
[37] Thanks to an anonymous reviewer for encouraging me to consider this implication.
[38] The discussion of psychopaths in relation to the internalism/externalism debate is extensive. In addition to Prinz, the following authors consider the possibility that psychopaths constitute real-life counterexamples to internalism: Kennett (2006), Matthews (2014), Maibom (2018), Nichols (2002), Smith (1994), and Sinnott-Armstrong (2014).

However, the inability of psychopaths to make moral judgments has been recently called into doubt. In a meta-analysis of research on psychopathic moral decision-making, Larsen et al. state that they 'found no empirical support for common perceptions of clinicians and laypeople that psychopaths are remorseless, unempathetic, and/or otherwise morally incapable' (2020, 10).[39] In terms of the position argued for here, these findings are significant in this respect: If moral judgments do not form a uniform kind, determining whether someone has a sufficient grasp of the use of moral concepts should become rather difficult to discern, and that seems to be just what Larsen et al. have found. If what we refer to as 'moral judgments' have different features in different contexts, this also seems to suggest a way forward: to determine whether an individual has a defective conception of morality, we would have to expose them to a whole host of moral decision-making contexts, and they may well be proficient in some areas but not others. This suggests an avenue for further research in this area.

Acknowledgements

Thanks to Rasmus Rosenberg Larsen and Alex King for their helpful feedback on this manuscript.

REFERENCES

Aharoni, Eyal, Walter Sinnott-Armstrong, and Kent Kiehl. 2012. 'Can Psychopathic Offenders Discern Moral Wrongs? A New Look at the Moral/Conventional Distinction'. *Journal of Abnormal Psychology* 121 (2): 484-497. https://doi.org/10.1037/a0024796.

Beebe, James, and David Sackris. 2016. 'Moral Objectivism Across the Lifespan'. *Philosophical Psychology* 29 (6): 912-929. https://doi.org/10.1080/09515089.2016.1174843.

Bedke, Matthew. 2009. 'Moral Judgment Purposivism: Saving Internalism from Amoralism'. *Philosophical Studies* 144: 189–209. https://doi.org/10.1007/s11098-008-9205-5.

Björnsson, Gunnar. 2002. 'How Emotivism Survives Immoralists, Irrationality, and Depression'. *Southern Journal of Philosophy* 40 (3): 327–344. https://doi.org/10.1111/j.2041-6962.2002.tb01905.x.

[39] Others have reached a similar conclusion: see Aharoni et al. (2012), Borg and Sinnott-Armstrong (2013), and Marshall et al. (2018).

Blackburn, Simon. 1984. *Spreading the Word*. Oxford: Oxford University Press.

Blasi, Augusto. 2001. 'Emotions and Moral Motivation'. *Journal for the Theory of Social Behavior* 29 (1): 1-19. https://doi.org/10.1111/1468-5914.00088.

Brink, David. 1986. 'Externalist Moral Realism'. *Southern Journal of Philosophy* 24 (1): 23-41. https://doi.org/10.1111/j.2041-6962.1986.tb01594.x.

Bromwich, Danielle. 2013. 'Motivational Internalism and the Challenge of Amoralism'. *European Journal of Philosophy* 24 (2): 452-471. https://doi.org/10.1111/ejop.12053.

Buckwalter, Wesley and John Turri. 2017. 'In the Thick of Moral Motivation'. *Review of the Philosophy and Psychology* 8: 433-453. https://doi.org/10.1007/s13164-016-0306-3.

Cameron, Daryl, Kristen Lindquist, and Kurt Gray. 2015. 'A Constructivist Review of Morality and Emotions: No Evidence for Specific Links between Moral Content and Discrete Emotions'. *Personality and Social Psychology Review* 19(4): 371-394. https://doi.org/10.1177/1088868314566683

Carroll, Noël. 2015. 'Defending the Content Approach to Aesthetic Experience'. *Metaphilosophy* 46 (2): 171-188. https://doi.org/10.1111/meta.12131.

Copp, David. 1995. 'Moral obligation and Moral Motivation'. *Canadian Journal of Philosophy* 25: 187-219. https://doi.org/10.1080/00455091.1995.10717438.

Damasio, Antonio and Gary Van Hoesen. 1983. 'Emotional disturbances associated with focal lesions of the limbic frontal lobe'. In *Neuropsychology of Human Emotion*, edited by K.M. Heilman and P. Satz, 85-100. New York: Guilford Press.

Dreier, James. 1990. 'Internalism and Speaker Relativism'. *Ethics* 101 (1): 6-26.

Francén, Ragnor. 2010. 'Moral Motivation Pluralism'. *Journal of Ethics* 14: 117-148. https://doi.org/10.1007/s10892-010-9074-y.

Gill, Michael. 2009. 'Indeterminacy and Variability in Meta-Ethics'. *Philosophical Studies* 145 (2): 215-234. https://doi.org/10.1007/s11098-008-9220-6.

Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People are divided by Politics and Religion*. New York: Pantheon Books.

Hare, Richard. 1952. *The Language of Morals*. Oxford: Clarendon Press.

Harman, Gilbert. 1977. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.

Hasting, Janna, Werner Ceusters, Barry Smith, and Kevin Mulligan. 2011. 'Dispositions and Processes in the Emotion Ontology'. *ICBO: International Conference on Biomedical Ontology*, 71-78.

Hume, David. 1896. *A Treatise of Human Nature*. Edited by L.A. Selby-Bigge. Oxford: Clarendon Press.

Goodwin, Geoffrey, and John Darley. 2008. 'The Psychology of Meta-Ethics: Exploring Objectivism'. *Cognition* 106: 1339–1366. https://doi.org/10.1016/j.cognition.2007.06.007.

Kennett, Jeanette. 2006. 'Do Psychopaths Really Threaten Moral Rationalism?' *Philosophical Explorations* 9 (1): 69-82. https://doi.org/10.1080/13869790500492524.

King, Alex. 2018. 'The Amoralist and the Anaesthetic'. *Pacific Philosophical Quarterly* 99 (4): 632-663. https://doi.org/10.1111/papq.12225.

Korsgaard, Christine. 1986. 'Skepticism about Practical Reason'. *Journal of Philosophy* 83 (1): 5-25. https://doi.org/10.2307/2026464.

Korsgaard, Christine. 1996. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

Kripke, Saul. 1980. *Naming and Necessity*. Cambridge: Harvard University Press.

Kristoff, Nicholas. 2015. 'The Trader Who Donates Half His Pay'. *The New York Times*. Accessed September 22, 2021. https://www.nytimes.com/2015/04/05/opinion/sunday/nicholas-kristof-the-trader-who-donates-half-his-pay.html

Kumar, Victor. 2015. 'Moral Judgment as a Natural Kind'. *Philosophical Studies* 172: 2887-2910. https://doi.org/10.1007/s11098-015-0448-7.

Larsen, Ramus Rosenberg, Jarkko Jalava, and Stephanie Griffiths. 2020. 'Are Psychopathy Checklist (PCL) Psychopaths Dangerous, Untreatable, and Without Conscience? A Systematic Review of the Empirical Evidence'. *Psychology, Public Policy, and Law* 26 (3): 297-311. http://dx.doi.org/10.1037/law0000239.

Maibom, Heidi. 2018. 'What Can Philosophers Learn from Psychopathy?' *European Journal of Analytic Philosophy* 14 (1): 63-78. https://doi.org/10.31820/ejap.14.1.4.

Marshall, Julia, Ashley Watts, and Scott Lilienfeld. 2018. 'Do Psychopathic Individuals Possess a Misaligned Moral Compass? A Meta-Analytic Examination of Psychopathy's Relations with Moral Judgment'. *Personality Disorders: Theory, Research, and Treatment* 9 (1): 40-50. https://doi.org/10.1037/per0000226.

Mason, Elinor. 2008. 'An Argument against Motivational Internalism'. *Proceedings of the Aristotelian Society* 108 (2): 135-156. https://doi.org/10.1111/j.1467-9264.2008.00240.x.

Matthews, Eric. 2014. 'Psychopathy and Moral Rationality'. In *Being Amoral: Psychopathy and Moral Incapacity*, edited by Thomas Schramme, 71-90. Cambridge: MIT Press.

Nichols, Shaun. 2002. How Psychopaths Threaten Moral Rationalism: Is it Irrational to be Amoral? *The Monist* 85 (2) 285-303.

Nichols, Shaun. 2004. *Sentimental Rules*: *On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.

Nichols, Shaun, and Michael Gill. 2008. 'Sentimentalist Pluralism: Moral Psychology and Philosophical Ethics'. *Philosophical Issues* 18 (1): 143–163. https://doi.org/10.1111/j.1533-6077.2008.00142.x.

Prinz, Jesse. 2007. *The Emotional Construction of Morals*. New York: Oxford University Press.

Richardson, Henry. 2018. Moral Reasoning. *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Accessed September 22, 2021. https://plato.stanford.edu/archives/fall2018/entries/reasoning-moral/.

Rosati, Connie. 2016. Moral Motivation. *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. Accessed September 22, 2021. https://plato.stanford.edu/archives/win2016/entries/moral-motivation/.

Sackris, David and James Beebe. 2014. 'Is Justification Necessary for Knowledge?' In *Advances in Experimental Philosophy*, edited by James Beebe, 175-192. London: Bloomsbury.

Sackris, David and Rasmus Rosenberg Larsen. 2020. 'A Consideration of Carroll's Content Theory'. *Journal of Value Inquiry* 54: 245-255. https://doi.org/10.1007/s10790-019-09693-6.

Sadler, Brooke. 2003. 'The Possibility of Amoralism: A Defense against Internalism'. *Philosophy* 78 (1): 63-78. https://doi.org/10.1017/S0031819103000044.

Setiya, Kieran. 2004. 'Against Internalism'. *Noûs* 38 (2): 266-298. https://doi.org/10.1111/j.1468-0068.2004.00470.x.

Shafer-Landau, Russ. 2000. 'A Defense of Motivational Externalism'. *Philosophical Studies* 97 (3): 267-291. https://doi.org/10.1023/A:1018609130376.

Shafer-Landau, Russ. 2015. *The Fundamentals of Ethics*. Oxford: Oxford University Press.

Sinnott-Armstrong, Walter. 2014. 'Do Psychopaths Refute Internalism?' In *Being Amoral: Psychopathy and Moral Incapacity*, edited by Thomas Schramme, 187-208. Cambridge: MIT Press.

Sinnott-Armstrong, Walter and Thalia Wheatley. 2012. 'The Disunity of Moral Judgment and why it Matters in Philosophy'. *The Monist* 95 (3): 355-377. https://doi.org/10.5840/monist201295319.

Sinnott-Armstrong, Walter and Thalia Wheatley. 2014. 'Are Moral Judgments Unified?' *Philosophical Psychology* 27 (4): 451-474. https://doi.org/10.1080/09515089.2012.736075.

Singer, Peter. 1972. 'Famine, Affluence, and Morality. *Philosophy and Public Affairs* 1 (3): 229-243.

Slote, Michael. 2010. *Moral Sentimentalism*. Oxford: Oxford University Press.

Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell Publishing.

Stangor, Charles, and Jennifer Walinga. 2014. *Introduction to Psychology – 1st Canadian Edition*. Victoria, B.C.: BCcampus. https://opentextbc.ca/introductiontopsychology/.

Stich, Steven. 2006. 'Is Morality an Elegant Machine or a Kludge?' *Journal of Cognition and Culture* 6 (1-2): 181-189. https://doi.org/10.1163/156853706776931349.

Stocker, Michael. 1979. 'Desiring the Bad: An Essay in Moral Psychology'. *Journal of Philosophy* 76 (1): 738-753. https://doi.org/10.2307/2025856.

Svavarsdottir, Sigrun. 1999. 'Moral Cognitivism and Motivation'. *The Philosophical Review*: 108 (2): 161-219. https://doi.org/10.2307/2998300.

Tresan, Jon. 2006. 'De Dicto Internalist Cognitivism'. *Noûs* 40 (1): 143-165. https://doi.org/10.1111/j.0029-4624.2006.00604.x.

Tresan, Jon. 2009. 'The Challenge of Communal Internalism'. *The Journal of Value Inquiry* 43: 179-199. https://doi.org/10.1007/s10790-008-9141-9.

EuJAP | Vol. 17 | No. 2 | 2021
UDC: 164:165.19
https://doi.org/10.31820/ejap.17.2.2

LOGICAL RELATIVISM THROUGH LOGICAL CONTEXTS

Jonas R. Becker Arenhart[1,2]

[1] Federal University of Santa Catarina
[2] Federal University of Maranhão

Original scientific article – Received: 06/07/2021 Accepted: 12/10/2021

ABSTRACT

We advance an approach to logical contexts that grounds the claim that logic is a local matter: distinct contexts require distinct logics. The approach results from a concern about context individuation, and holds that a logic may be constitutive of a context or domain of application. We add a naturalistic component: distinct domains are more than mere technical curiosities; as intuitionistic mathematics testifies, some of the distinct forms of inference in different domains are actively pursued as legitimate fields of research in current mathematics, so, unless one is willing to revise the current scientific practice, generalism must go. The approach is advanced by discussing some tenets of a similar argument advanced by Shapiro, in the context of logic as models approach. In order to make our view more appealing, we reformulate a version of logic as models approach following naturalistic lines, and bring logic closer to the use of models in science.

Keywords: *Classical logic; intuitionistic logic; relativism; logic as models; context constitution*

Correspondence: jonas.becker2@gmail.com

1. Introduction

Logical generalism, as the name suggests, is the thesis that logic is general. This is ambiguous in the same measure as the term 'logic' is: on the one hand, it may denote 'logic' as a discipline, on the other, it may denote 'logic' as a specific system of logic. As Shapiro notes,

> Moreover, logic is ubiquitous. [...] there is a longstanding view, with a stellar pedigree, that logical consequence is topic neutral; it applies everywhere. Even if that is challenged, [...] it remains that every coherent perspective—every language, every form of life, every context—has a logic. (Shapiro 2014, 165)

Here, we shall not discuss whether logic, taken *as a science*, is general, or universal, or ubiquitous. Rather, we shall focus on the claim that logic is general when one considers distinct *systems of logic* attempting to capture the validity of inferences in natural language (the so-called 'canonical application'; see Priest 2006, 196-197). Our claim, again, is that distinct systems are required and legitimate for distinct contexts, even when the field of application concerns inferences in natural language.

One could complain about that way of framing the problem, which emphasizes the role of distinct *systems* of logic. It could be said that validity in distinct systems should not be confused with validity *per se* or validity *tout court;* certainly, the claim could go, distinct systems characterize distinct notions of the consequence relation, but that is not what is at stake in philosophical debates. Rather, what is being disputed is whether distinct notions of validity are legitimate, or correct. That is, the question is whether there are, out in the wild, different notions of validity that require distinct systems to be characterized, and of which these systems are said to give a correct/incorrect description (depending on the case).

Now, although one *could* advance such an objection, our discussion will not presuppose that there is such a thing as validity *per se*, or validity *tout court*. However, as we shall see, there is a sense to be made of claims of distinct systems being correct for distinct contexts. Basically, the notion of correctness, as our proposal will characterize it, does not require correspondence of a theoretically described notion of validity with an independently existing notion of validity, out in the wild (this will be discussed latter). Furthermore, given that we can only characterize the distinct notions of validity that are in dispute in terms of some logical theory, using the logical apparatuses furnished by distinct systems, we

shall keep with our talk of distinct systems, and not talk in terms of validity *per se*. Certainly, much more could be said about the idea that logical correctness is related to the correspondence of a system of logic with an intuitive or pre-theoretical notion of validity, but this is enough as a warning to begin with. We shall avoid this notion in our discussion, given that a proper treatment of this problem would require a different route.

With those points out of the way, let us proceed. Our next task is to attribute a more precise meaning to the idea that logical theories, or logical systems, can be claimed to be general or local. There are some claims advanced to that purpose, although none of them provides a precise characterization that could be adopted as an official definition. As Routley (1980, 83) has advanced the claim of generalism, approvingly, what lies behind the generalist thesis is a worry about the scope of logic, the fact that "[l]ogic is not merely a local matter, and should, insofar as it is correct, apply universally." Notice that this connects correction and generality. The opposite of generalism, a form of localism, may be understood then as the claim that logic is a local matter; a logic may be correct only locally.

Some opponents of logical generalism go in the same direction when it comes to characterizing the core of the generalist thesis. Wyatt and Payette (2021, 4813), for example, characterize generalism as the claim according to which "logical systems and logical laws must have universal application". Dicher (Forthcoming, 2), also not a defender of generalism, characterizes generalism as consisting of the view that, on what concerns logic, "there are no exceptions to its laws, which apply across every domain of inquiry, irrespective of the particular features of that domain". Again, the most important feature of generalism concerns the claim that logic meets no borders; a system must have its inferences and validities applying in *every context*. Logic would be *local*, then, if its laws would have local applicability or validity, if distinct systems were required to account for distinct domains. Hjortland (2013, 356) frames the localist claim in terms of the existence of at least two *domains of discourse* for which correct deductive reasoning requires distinct logics.

What these characterizations have in common, together with the discussion on generalism is, the idea of a *context*, or a *domain*, along with the claim that logic must be correct, or applied properly, irrespective of the context, or domain. Generalism involves the claim that a logical theory applies in every context or domain, it is insensitive to the demands of each particular domain it may meet. The *individuation of a domain*, then, is granted independently of the underlying logic; that is, according to generalism, in specifying a domain or context, the underlying logic is taken for granted (given that it is universal, context-independent), and each domain builds

over it with its specific features (the context-depend ones). According to generalism, what is common to all domains or contexts is the underlying logic.

Notice that one may regard that only one system of logic fits the bill of being general enough, of accounting for every domain, resulting in a *monist* position about logic, or else one may hold that distinct systems of logic are all equally successful in being general as required, resulting in *pluralism* about logic. The distinction general/local does not collapse into the distinction monism/pluralism, although it is much more common to find monists among generalists than pluralists. For localists, those that hold that distinct logics may be correct or appropriate for distinct contexts, the same distinction applies. Given a context, one may believe that only one system of logic is correct for that context (*local monism*, a position defended by da Costa 1997), or that a whole family of distinct systems may be equally correct for that context (*local pluralism*, a position defended by Bueno 2002, for instance).

The terminology thus introduced requires that we distinguish between the pair local/general on the one hand, and the pair one/many on the other, when it comes to logic.[1] Their combinations give rise to the current spectrum of traditional positions: logical monism and logical pluralism, as traditionally understood, are generalist theses, holding that there is one and that there are many correct logics, respectively. Relativism or localism is the thesis that logic is local, and the question remains open as to whether there are many distinct logics for one context, or only one for each context.

In this paper, we shall focus on the general/local divide, leaving the issue of one/many for another occasion. Our plan is to elaborate over already existing proposals for logical relativism, and we do so by putting logic in a naturalistic setting in two related senses. First of all, naturalism is understood as the methodological claim that there is no first philosophy to judge science, with logic and mathematics understood as part of science. Second, the approach advanced here is naturalist also in the way that the 'logic as models' approach is framed, requiring that models be understood in closer connection to the workings of models in science; more specifically, we shall suggest that the understanding of models in science according to the view called 'models as epistemic tools', as developed by Knuuttila and Boon (2011), can be fruitfully adapted to the case of logic. This will provide us the appropriate understanding of 'context' required to motivate localism in logic. As we shall see, logical generalism is not

[1] This clearly complements the distinction advanced by Haack (1978, 223) and, following Haack, by Hjortland (2013, 356-357).

motivated, *if* an account of logic that is more naturalistic is adopted, *and* when the notion of a context is properly understood in connection to scientific modeling. In order to motivate our proposal, we shall briefly discuss a related argument advanced by Shapiro (2014), who also defends the logic as models approach. We shall use what appear to be some *tensions* in Shapiro's approach to suggest an alternative account that not only overcomes the difficulties, but also presents some virtues that recommend it as a better option for the friends of the logic as models approach. As a kind of bonus, we hope, the resulting combination of naturalism and 'logic as models', as developed here, can be used to articulate a version of logical anti-exceptionalism; according to the latter, logic is continuous with empirical science in many respects (see Hjortland 2017). Perhaps, the view defended here does contribute to substantiate this claim, although we shall not develop it here.

Perhaps one more word on the pluralism/monism divide is in order. Typically, this is directly connected with the question of whether one or many logics are correct, and the problem of the correction of a logic is a substantial one, concerning connection of the formal systems with extra-systematic considerations about validity (see Haack 1978, chapter 2). As we shall propose in the paper, due to the kind of approach to logical contexts we advance, the 'correct' logic for a context becomes somehow an *a priori* issue, not open for substantial dispute (of course, the topic is developed in the paper). The locus of dispute, due to the naturalistic approach to the epistemology of logic and theory choice shifts, then, to the dispute on whether it is one or many logics that are currently required by the scientific community in its investigative practice. In this sense, the debate resembles the 'monism versus pluralism' debate, but the locus of importance is shifted, given that the issue of correction of a logic is mostly deflated. Developing this difference in depth would require a different paper, so that we just leave this as reminder for the reader. For those willing to keep the terminology, the view defended here would be classified as a form of local pluralism, although, again, the 'pluralism versus monism' debate is typically framed in terms that are considerably different from the one presented in the current paper.

The rest of the paper is structured as follows. In the next section, we advance Shapiro's argument against generalism, in the context of his approach to logic as models. In section 3, we present what may look like some difficulties for the strategy employed by Shapiro, and in particular, his understanding of the role of logic when it is considered under the logic as models approach. Our own suggestion arises as a solution to overcome the mentioned difficulties, and comes from a twist to Shapiro's perspectives. We argue that it combines perfectly with a more science-

friendly approach to models in science in general, known in the philosophical literature as the 'logic as epistemic tools' approach. We conclude in section 5.

2. Shapiro's approach

We start by presenting Shapiro's approach against logical generalism. As we understand it, it requires Shapiro's account of logic as *per* the *logic as a model* approach as a starting point. Basically, Shapiro holds that systems of logic are to be understood as models of inferences in natural language, in what is regarded as the same sense that 'model' is understood and employed in the sciences. This holds explicitly for formal languages:

> I propose that a formal language is a *mathematical model* of a natural language in roughly the same sense as, say, a collection of point masses is a model of a system of physical objects, and a Turing machine is a mathematical model of a person following an algorithm, or perhaps a computing device. In other words, a formal language displays certain features of natural languages, while ignoring, simplifying, or idealizing other features [...]. (Shapiro 2014, 46)

Besides language, the modeling account also deals with the notion of logical consequence. The similarities of use of mathematical models in logic with the understanding of how models are used in other areas of investigation results in the question of the correctness of systems of logic being largely relative to our specific purposes, and to the accompanying claim that their success should be evaluated accordingly. As Shapiro claims:

> With mathematical models, which features one focuses on, which are idealized, and which are ignored, depends on the purposes at hand, on why one is developing a model in the first place. Here, of course, our goal is to shed light on the relation (or relations) of logical consequence, and perhaps the norms for deductive reasoning and regulating beliefs to maintain consistency. So, presumably, in developing a logic-model, we should focus on and idealize those features of natural language that bear on deductive reasoning, or on regulating our beliefs for consistency, whatever those features may be. (Shapiro 2014, 47)

This opens the door for arguing that distinct logics may be appropriate for distinct fields or contexts, given that we may have different purposes in different contexts. Indeed, this is the crucial ingredient for the kind of argument that Shapiro will advance for the relative character of logic. In order to do so, Shapiro couples this view on logic with a form of Hilbertianism on the philosophy of mathematics. Basically, this is an update on Hilbert's *motto* according to which, roughly speaking, *consistency implies existence*. That is, a mathematical structure that is consistent implicitly defines the entities it deals with, just as in Euclidean geometry the notion of point is defined implicitly by the geometrical axioms for 'point'. Given, however, that consistency is a matter of which logic one uses, and that distinct mathematical structures will end up being inconsistent when certain logics are adopted as their underlying logic, the result is that given our purpose of preserving consistency, perhaps distinct logics are required to account for the perceived consistency of distinct mathematical structures.

Notice how the dialectics to ensure relativism goes. First, it is assumed that a kind of mathematical pluralism holds. This means that distinct kinds of mathematical structures are legitimate due to their consistency and to their actual interest for mathematicians. As a second step, given this plurality, we may inquire over which logics are required to make such mathematical theories consistent, or, in other words, for which are the appropriate underlying logics of such theories. If it happens that distinct mathematical theories require distinct logics, then we are justified in adopting a form of localism about logic, that is, that distinct logics apply in distinct domains or contexts (Shapiro calls this *relativism*). Here, the logic as models approach is playing a major role: given the diversity of mathematical structures taken as legitimate as a kind of point of departure, or as a kind of 'neutral' data that appears to be independent of the issue of which logic or logics are appropriate, we idealize the inferential practices of existing distinct mathematical theories in order to comply with the demands of consistency in each case. The result, as claimed, is that distinct theories will end up requiring distinct logics if their internal consistency is to be preserved.

The case for the general argument is made with an illustration employing intuitionistic mathematics. As it is well-known, intuitionistic theories conflict with classical mathematics, and this conflict concerns the inferences available in each case (see Shapiro 2014, chapter 3, for specific examples concerning intuitionistic theories: Peano arithmetic using intuitionistic logic plus Church's thesis (PA+CT), the intuitionistic analysis, and smooth infinitesimal analysis (SIA)). Let's focus on the

simple case of intuitionistic analysis. In classical analysis, developed using classical logic, it is possible to define real functions that are not continuous (this is widely known, of course). Intuitionistic analysis does not vindicate such a simple fact, and this surprises students of classical mathematics when they hear of it for the first time. How can that be? This is a direct result of the theory of real numbers adopted by intuitionists.

One of the sources for the difference is to be found in the very concept of real number in intuitionistic analysis. Indeed, the intuitionist may consider a real number as an equivalence class of Cauchy sequences of rational numbers, just as a classical mathematician does. However, a sequence, in this context, is a choice sequence, it is only *potentially infinite*, never complete (intuitionists do not accept complete infinities, remember). Recall that a sequence s of rational numbers is Cauchy if for every rational number $\varepsilon > 0$, there is a natural number N, so that for every natural numbers $m > N$ and $n > N$, $|s(n) – s(m)| < \varepsilon$ (intuitively, the terms of the sequence s may be seen as approaching each other, as the function picks as arguments numbers standing after a given N in the usual order). For an intuitionist, given any ε, if a sequence is Cauchy, *one must be able to compute* the N after which the members of the sequence are within the ε given. Over such a view of real numbers, we have:

> *Brouwer's theorem*: all real functions defined over a closed interval are uniformly continuous.

The details of the proof of the theorem need not concern us here. What is more relevant is that the result conflicts with classical analysis, and that Shapiro makes use of this fact to argue for the requirement of distinct logics for distinct mathematical structures. He argues that if we add the law of excluded middle to intuitionistic analysis, we are able to define functions that are not continuous. In this sense, then, Brouwer's theorem holds only in the presence of restrictions to classical logic; it requires intuitionistic logic. Classical logic is not *consistent* with it.

As a result, given that intuitionistic analysis is taken as a legitimate kind of mathematical structure (the *initial data*, recall) deserving to be developed and investigated, a kind of relativism about logic arises, due to the fact that distinct legitimate mathematical theories require distinct logics to be consistent (notice also the naturalistic bent, bringing mathematical practice to guide theory acceptability, rather than philosophical claims). The result is a restriction on the applicability of classical logic, as well as of intuitionistic logic. As Shapiro puts it:

> conceding that the law of excluded middle, and thus classical logic, is not universally valid. That is, classical logic is not correct in all discourses, about all subject matters, etc. The intuitionist is right about that much. (Shapiro 2014, 82)

That is, the conclusion is precisely a denial of generalism (given that Shapiro admits that excluded middle holds in classical analysis). Given that intuitionistic logic is required to account for part of that practice in intuitionistic analysis, it seems that intuitionistic logic is legitimate as the underlying logic of a domain of investigation. Of course, given that one is also assuming that classical structures require classical logic, then, distinct domains or contexts require distinct logics. This would justify rejection of the version of generalism we are concerned with.

3. Some tensions for Shapiro's account

Although, as it will become clear, we are in agreement with the main conclusion established by Shapiro, we still seem to find some sources of tension that must be acknowledged in Shapiro's path leading from Hilbertianism about mathematics to logical localism. In this section we shall bring some of them to the fore. Avoiding such tensions is the major goal of the approach we shall advance in the next section.

The first source of concern is related to the requirement of consistency preservation as a sign for the appropriateness of a logic for a given context. That is, to recognize that a logic is appropriate for a given context, one is required to check whether that logic preserves or grants the context's consistency (relative to that very same logic). Although it seems quite reasonable in the context of mathematical theories, the worry is that it may lead one to the wrong kind of account of the underlying logic in some quite interesting cases. There are historically well-known cases, such as Frege's *Grundgesetze*, where choice of the underlying logic is out of the question, but still, the system is not consistent. Still, despite its inconsistency and triviality, the system is not without logical interest. Also, for a more recent episode, da Costa's original formulation of his paraconsistent version of the set theory NF (New Foundations) was established as trivial, although there was no question of the choice of a logic (see da Costa 1986). Again, although the logic chosen by da Costa was not properly ensuring consistency (in this case, non-triviality), the system was clearly interesting, and had some important lessons to teach on the nature of paraconsistent set theory.

As a result, the requirement of consistency does not seem to provide, in some cases, at least, the best help when it comes to connecting systems of logic with specific contexts. Some legitimate mathematical contexts can be said to have well determined logics that clearly violate the requirement. In other words, some contexts come with a logic, explicitly formulated, that violate the requirement of consistency. In these cases, it does not seem appropriate to hold that the logic leading to inconsistency/triviality did not contribute to the individuation of each context. They did, but it ended up being the case that the systems were inconsistent/trivial. The very idea that one can attempt to fix Frege's system, as neo-logicists do, or that da Costa could fix his system, only makes sense if we accept that the original system is and remain inconsistent/trivial. Accepting that the underlying logics help us characterize and individuate a context indicates that changing the underlying logic will result in a different theory (more on this soon).

Perhaps one could object to this point in the following way: the cases brought here do not cause a problem to Shapiro, given that Shapiro is only concerned with what are considered legitimate mathematical structures, that is, consistent contexts really investigated by the mathematical community. Being inconsistent, Frege's G*rundgesteze* is not legitimate; being trivial, da Costa's theory is not legitimate, and Shapiro would have nothing to do with them. However, it seems to us that this would limit the interest of Shapiro's approach, missing interesting facts about the comings and goings of mathematical structures. For example, consider Cantor's naive set theory. With the discovery of Russell's paradox, the theory was fixed in a plurality of alternative ways, and interest in it did not disappear due to inconsistency. So, in a sense, it may happen that some theories are individuated with logics that do not grant them consistency, or non triviality. However, that does not mean that such theories cannot be interesting from a mathematical point of view. Rather, people try to keep some of the results of the theory either by changing the logic, or by changing some of the axioms specific to the theory (in both cases, with new contexts arising). That is, some theories may be on the radar of mathematicians even if they are inconsistent, and the search for a consistent version may be even a part of the pursuit of such mathematicians.

The second perceived source of tension concerns the very idea of a context. Shapiro has offered the following characterization of a context:

> I propose that each "context" includes a specific mathematical theory or structure. It would be the mathematical theory being advanced at any given time by a mathematician or a group of mathematicians. In line with the foregoing eclectic orientation,

> each such context has a specific logic: classical logic for the classical theories, intuitionistic logic for the intuitionistic ones, etc. Sometimes we will just think of a logic alone as a context, if the ambient mathematical theory is not in focus or does not matter. (Shapiro 2014, 89)

However, in inferring from mathematical pluralism to logical localism, one must acknowledge that the adoption of distinct logics is *a result* of distinct contexts being already considered as legitimate, which in this case are the *distinct mathematical theories* currently investigated by the mathematical community. But distinct mathematical theories (which are playing the role of the contexts, here) seem to be characterized as incorporating a logic beforehand. That is, *a logic is part of what constitutes and individuates a context*. Under these conditions, it seems implausible to think that we can have a context (when this is a mathematical theory) individuated independently of a logic, only afterwards extracting from it a logic. Alternatively, we could proceed as Shapiro implies, seeing logics as models, and attempting to model the inference patterns of the context in each case where the context is legitimate in the eyes of the mathematicians. That is, logics are there to begin with, characterizing the context, but also, must be extracted from the context. So, the dilemma may be put as follows: on the one hand, the logic as models approach requires that we somehow idealize from given practices of inferences, generating a set of inferential patterns considered appropriate for the goal of preserving consistency in that context. On the other hand, a context is specified with the help of a logic. But then, we seem to be in trouble: logic must be already there to define a context, and also, be extracted from a context by the modeling procedure. It seems that we cannot have it both ways.

Given that this issue is of central importance for our own argument against logical generalism, let us check what is going on in more detail. To motivate the failure of generalism, one must argue that distinct logics are required for distinct contexts. Shapiro attempts to grant that fact by starting with distinct mathematical theories that are playing the role of the contexts and provide a kind of neutral data on the issue of which logic is appropriate. Given these contexts, he proposes to somehow extract, by means of the modeling approach, the required logics that account for their consistency, by checking which system preserves the consistency of each mathematical theory. This would make a case for distinct logics in distinct contexts that is not question begging, and that confers credibility to the view, given that the distinct contexts one started with are scientifically respectable. However, when it comes to defining a context, systems of logic already play a role in their individuation. If this is really so, as it is suggested by the characterization of a context, then, one cannot really have a fully

convincing argument for the failure of generalism, given that such distinct logics were admitted as legitimate right from the start, with the claim that distinct mathematical theories constitute distinct contexts, and that such theories come with a logic.

We can make the point quite forcefully considering Brouwer's theorem, mentioned earlier. The fact that it is proved in intuitionistically acceptable ways already points to the need of identifying some logical resources, and that the law of excluded middle is not one of them. This should give us pause to think that perhaps intuitionistic logic (or something quite similar) must be available in the background beforehand, otherwise the context would be developed very differently. At least when it comes to mathematical theories, it is quite difficult to think that one could have some inferential practices in developing the theory that latter, under closer analysis, turn out to be intuitionistic logic, without consciously applying them in order to develop the mathematical context to begin with. In this sense, logic is required to characterize the context. However, as we have seen, the move by Shapiro also seems to require that logic is established after the context is available, by some kind of modeling activity. There lies the tension.

One could hold that the tension is illusory.[2] In fact, it can be argued that every context comes with an underlying logic L, but then, with the development of the theory, still inquire whether the theory is really consistent with L. It may turn out that it is, and that L is the best model for the kind of inference used in this context, or it may turn out that a distinct system of logic may be more adequate, resulting in the case that the logic discovered after the modeling process is applied, L', is different from L, but still, more appropriate than L. This, it could be claimed, could make the use of an underlying logic to individuate a context compatible with the use of the logic as models approach to obtain a logic from the context, dispelling the kind of trouble that we have attempted to point out. But notice that this objection cannot dispel the worries we have raised. Given that a logic L is presented as the underlying logic for a context, it would be odd, to say the least, to discover, afterwards, that we did not properly infer according to it, so that the modeling process of our inferences ended up delivering a different logic. Why start with L, then, if we are not required to infer according to it? If it happens that the logic L leads us to triviality, such as in Frege's *Grundgesetze* case, then, of course, we can only discover that by really using the logic L. This allows us to fix the context, in case it is of mathematical interest, originating new contexts (as discussed earlier also in the case of Cantor's theory and paraconsistent set theory).

[2] Again, I owe the objection to an anonymous referee, to whom I would like to thank.

As a result, we keep the claim that there is a tension in locating logic in the context to begin with, and also attempting to extract it from the context by use of the modeling approach. There are basically two views on the identity of context playing a major role here. On the one hand, a context is individuated by its underlying logic, so that the logic comes with the context. On the other hand, a context is given independently of a logic, so that the logic appropriate for the context is identified by a process of modeling of the inferential practices of users, restricted also to the demands of consistency. The argument from mathematical pluralism to logical localism depends on the latter view, it seems, because the plan is to infer the diversity of logics from the diversity of mathematics, by observing the demand of consistency.

One can avoid the difficulty either by attempting to define 'context' without the use of logic as a constitutive component, which seems difficult in this circumstance, or by providing for another account leading from distinct contexts to the acceptability or correctness of distinct logics in such contexts. Our proposal consists in following the second route, and we shall see logics as contributing to the individuation of contexts in a more thoroughly naturalistic approach.

4. Inverting the perspective

In order to avoid the tension mentioned in the previous section, due to the very nature of a context, we shall acknowledge right from the start that systems of logic do act as (at least partially) constituting contexts. That is, in our view, a logical theory contributes actively to the individuation of a context; contrarily to what the generalist suggests, that a fixed logic is taken for granted, and that the specific contents of a context are added on the top of it, we allow that even a system of logic may be used to legitimately individuate a context. As we shall argue from now on, this has at least two main advantages: it avoids the problem of an apparent kind of circularity in justifying the use of a logic in a given context, and also the problem of deciding issues of the right logic for a context (without requiring that there is a notion of validity *per se*, as mentioned at the beginning of the paper). These issues are solved by the more flexible notion of context that we advance.

It should be recognized that when it comes to mathematical theories, at least, logics are indeed part of the characterization of their respective contexts. To begin with the motivation for such a characterization of context, and the claim that it leads to localism quite directly, notice that

this will already make a better sense of the currently developing literature of *inconsistent mathematics*. Even though inconsistent mathematics does not enjoy (at least for now) the same kind of wide acceptance of intuitionistic and classical mathematics, it is a field that has been growing in recent years. Consider the following definition of inconsistent mathematics:

> Inconsistent mathematics is the study of the mathematical theories that result when classical mathematical axioms are asserted within the framework of a (non-classical) logic which can tolerate the presence of a contradiction without turning every sentence into a theorem. (Mortensen 2017)

In other words, inconsistent mathematics are the mathematical theories developed over paraconsistent logics (see also Priest 2006, chapter 10 and the definition of inconsistent arithmetic). This defines a family of contexts in which paraconsistent logics are the correct logics, by *fiat*, as it were. Clearly, distinct kinds of paraconsistent structures require distinct kinds of paraconsistent logics, and the logic must be clearly specified right from the start. Now, if other contexts may be defined in the same way, and are considered legitimate by anyone in the dispute, then, there is a good case against the generalist.

Before we proceed, notice how, in the case of inconsistent mathematics, such systems of logic are allowing us to individuate the contexts in question; inconsistent mathematics is defined as employing paraconsistent logics to begin with. In an important sense, there is no paraconsistent mathematics as a kind of activity first, and afterwards, we go on looking for the inferential patterns that enable such mathematics (that make it 'consistent', meaning 'non-trivial' here). The direction suggested by Shapiro, of going from the mathematics to the logic, would hardly work here. Rather, without such logics there to begin with, there would not be a case for the existence and complete understanding of the identity of such contexts. In the case of mathematical contexts, the logics are assumed by default, and they are the correct logics for the specific contexts they help individuate to begin with. In the same sense, we suggest, it would be odd to have intuitionistic mathematicians, and classical mathematicians too, proving theorems, each in his or her own domain, and only afterwards looking for their specific inferential patterns, in order to investigate which logic is more suitable. The patterns codifying valid inferences are not there somewhat hidden, awaiting to be found by *a posteriori* modeling activity. Rather, they are set at the beginning, to individuate the context. The logics act as enabling the development of the kind of mathematics of which they are the underlying logics.

It is not the case, then, that distinct mathematical theories or structures make a case for the plausibility of the use of distinct logics; rather, distinct logics act to enable that distinct mathematics be developed. This solves the problem of determining the individuality of a context, avoiding what was perceived as a kind of circularity in the previous section. It puts the issue of the appropriate direction of the dependence of a context on logic on a clearer basis: the adoption of distinct logics is not a result of the acceptance of distinct mathematical structures as legitimate; rather, the distinct mathematical structures are a result of distinct approaches to logic, which act as a guide in the development of such mathematical structures. Although this may sound historically inaccurate in some cases, due to Brouwer's distrust of logic in general, there is a case to be made for it, even on what concerns intuitionism. Recall that although Brouwer did not develop a system of intuitionistic logic, his own approach to constructive mathematics originates in great part from his distrust of classical logic, and on restrictions to classical inference modes. In fact, in order to characterize constructive mathematics, in general, one needs to appeal to the kind of inferences, or logical behavior that is the basis of such contexts:

> Constructive mathematics is distinguished from its traditional counterpart, classical mathematics, by the strict interpretation of the phrase "there exists" as "we can construct". In order to work constructively, we need to re-interpret not only the existential quantifier but all the logical connectives and quantifiers as instructions on how to construct a proof of the statement involving these logical expressions. (Bridges and Palmgreen 2018)

In this sense, just as inconsistent mathematics is mathematics developed over paraconsistent logics, constructive mathematics requires a constructive understanding of the logical apparatus to begin with; the logic contributes to the identity of the context. And we may go even further, and consider classical mathematics, which was here even before something like classical logic was available, right? How can it be that classical logic acts as enabling it? Well, notice that the epithet 'classical' was applied to classical mathematics only after classical logic consolidated. Classical logic is a recent invention, and a distinction between 'classical' mathematics and other types of mathematics is only available after the consolidation of classical logic. So, in this sense, the distinct logics and inference patterns required for distinct contexts, in the case of mathematical theories, are part of the very definition of a context, and are correct for those contexts due to this very fact.

But now, given that the localist thesis is no longer inferred from a given neutral data, the plurality of mathematical theories, how do we grant that distinct logics are required for distinct contexts? Or, in other terms: how do we grant that distinct logics are required for distinct contexts, and that they are *correct* for them? We need to separate two distinct issues that are conflated in this kind of question. One way of looking at the question is concerned with the correctness of a logic for a specific context. This, at least in the case of some mathematical theories, is solved by the appropriate, and more refined, notion of a context that we advanced. A logic is already employed when it comes to characterizing a context, and is the appropriate logic for that context. After defining contexts like that, right from the start there is a logic that is doing the work of being the underlying logic. One could believe that this makes the issue of the correct logic rather uninteresting; in fact, this brings the disputes over the appropriate logic to a quick solution.[3] However, although this may be seen as deflating some of the disputes over the correct logic, which may be seen as a virtue by some, it also shifts the locus of interest to another question: which such systems are interesting, or worthwhile pursuing?

This is in fact the second question that is conflated with the previous one. It concerns the respectability, from a scientific point of view, of each such context that may be advanced for the consideration of the scientific community. Classical mathematics clearly has an upper hand here, given its long intellectual tradition and successful application to empirical sciences. But intuitionistic mathematics is also an institutionally recognized scientific research program. Anyone ruling one of such contexts out would be adopting a revisionist program of the philosophy of mathematics that does not account for the practice of the discipline in our days, and as such would have the burden of proof.

This may be put in the context of the Carnapian principle of tolerance. The principle requires that we allow distinct systems to be investigated, and not to discard them based on philosophical prejudice. However, tolerance is still not enough to grant scientific respectability and ensure wide adoption of such systems in research programs. Tolerance concerns the fact that each one is free to advance a framework as something worth of pursuit; this, by itself, does not grant that the system will be pursued. Only science, as an institution, determines which systems (understood here as mathematical structures) are worth of investigation.[4] Certainly, classical and intuitionistic mathematics pass this latter test. Given that each require

[3] Thanks to an anonymous referee for pointing that.
[4] Thus, logic and mathematics may also be seen as providing for research programs, in a Lakatosian sense, as suggested by Priest (1989).

a distinct logic, a form of logical localism in current mathematical practice seems to be vindicated. In this sense, then, generalism fails, because it cannot account for the current state of mathematics. One can be generalist only at the price of rejecting mathematical practice, which is possible, but not totally recommendable from a naturalistic point of view.

The, in a sense, we suggest a division of labor between the question of correctness of a logic for a given context and the question of what makes a system an interesting object of research. In the picture suggested, although the question of correctness becomes deflated, there is still an issue of whether the diversity of systems available can become an integrating part of current scientific enterprise. From the relativist point of view advanced, a plurality of systems is justified in the measure that they are part of such an ongoing enterprise. As Caret proposed:

> An honest naturalist simply takes mathematics as it stands and respects the autonomy of the discipline, rather than imposing outside ideas about how it 'should' be practiced. Who are we to police the bounds of mathematics because of some hangup about bivalence or truth-tables? (Caret 2021, 4964)

Such a practice recommends that some non-classical structures are currently part of the mathematical practice and this legitimates them. Notice that the issue of whether intuitionistic logic is correct for such practice is a prior issue (and here we differ from Caret); the point of relevance is accepting intuitionistic mathematics as part of the scientific enterprise. Again, this makes relativism interesting, the fact that it is anchored in the practice of science.

Let us contrast this approach again with Shapiro's strategy. While Shapiro uses the fact that our scientific community recognizes diverse mathematical structures as worthy of study and engagement as a starting point, which then leads to contexts and, from them, to distinct logics, we use logic as enabling the development of distinct mathematical theories, which, then, are acknowledged (or not) by our community as worthy of development (as fruitful research programs). That is, both approaches will have to appeal to the verdict of the scientific community on what concerns distinct mathematical structures and their scientific respectability as fruitful mathematical programs of research. However, while Shapiro uses this fact as a springboard to logical localism, attempting to ground the need for distinct logics in this fact, we use distinct logics to provide the very source of such distinct contexts. Scientific respectability comes after that, if it ever comes for some of the mathematical theories that are proposed. This describes perfectly well the situation of the inconsistent mathematics

program: this is clearly a program where it is known, beforehand, which logic is the underlying logic of the enterprise. What friends of paraconsistent mathematics claim is that such structures are also worth of investigation, that the mathematical community should also join the efforts of developing such structures, due to theoretical rewards to be expected. Whether the mathematical community will listen to the call, time will tell, but it is largely an issue concerning the practice of mathematics, not of choice of the appropriate logic.

There are many advantages in reversing the approach to contexts as we have done. First, we have a clearer identity condition for contexts; mathematical theories are not entities awaiting for a logic to be attributed to them; rather, they are endowed from the start with prescriptions for the correct inferences. Second, this solves by default the issue that distinct logics are required for distinct contexts, basically, because the logic is already an ingredient of the context. Third, it is compatible with a version of the tolerance principle in which distinct logics may be used (as they indeed are, as the case of paraconsistent mathematics attests) to advance different contexts, which are then developed in the hope that the community may somehow recognize their importance.

The approach is very logic-oriented; it makes use of the fact that the very notion of 'domain' gets broadened with the rise of non-classical logics, and with the recognition that logics themselves may be used or required to characterize contexts. This allows for distinct logics being used in distinct contexts by *fiat*, something that could not be imagined when such distinct logics were not available. So, the anti-generalist has a somehow direct case once distinct logics are present to constitute distinct domains. The point is that the easy case can become also epistemically respectable when such contexts are also scientifically relevant, and this is what happens with the intuitionistic mathematics, for instance. This is as far as a naturalist would demand of justification for the distinct logics, that they be really part of current science, and is compatible with tolerance with the development of alternative approaches, which then will look for their place in the scientific enterprise.

We can finish now with a short discussion about how the logic as models approach suits in the picture, once this new understanding of context is adopted. Recall that we have suggested that logics help individuate a domain, instead of first having a domain or context, and then looking for the logic. In this sense, recall, the proposal is quite logic-centered, in the sense that it allows that logics may contribute to inform, in a sense, the nature of a domain or context to which they are applied. This also means that a logic and a context are not independent entities, awaiting to be

matched. Rather, the logic somehow contributes to give a more specific shape and identity to the field of its own application. In more general terms, then, in the picture being proposed here, one needs an account of models that sees models as contributing actively to the character of the target they are intended to apply to. Typical accounts of models do not see models as having so much to offer on the way to individuate their targets or contexts of applications; rather, the typical accounts focus on the relation of models and targets, as two independent entities.

This situation reflects itself in the fact that most accounts of the role of models are still very much focused on the representation relation. The plan is that there are models on the one side, and targets on the other, and that knowledge about the target is obtained when the models are properly related to their targets. These accounts all recognize the role of abstraction, idealization and simplifications, but still, this is not enough to precisely account for the epistemic role that models play in our scientific activities:

> Apart from simplifications, approximations and idealizations, scientific modelling involves significant conceptual work, which covers such epistemic activities as discerning specific types of phenomena, conceptualizing 'non-directly observable' objects, properties, or processes, and bringing phenomena under specific types of 'non-empirical' theoretical principles or concepts. It is difficult to see how these conceptual activities would fit into the traditional representational picture. (Knuuttila and Boon 2011, 313)

That is, the traditional accounts (the 'representational picture' mentioned by Knuuttila and Boon) fall short of providing for a detailed enough picture of modeling. In particular, they fail to acknowledge the role of models in enabling the investigation of the target.

Luckily, there are proposals in the literature on the use of models in science that bring the required constitutive-enabling relation of the models to their targets to the center of the stage. Here, we shall propose that one may adapt the 'models as epistemic tools', advanced by Knuuttila and Boon (2011) to the case of logic, and get a result that is quite connected with the proposal we have been describing for the localist picture in logic. This account of models allows that a model play an active role in individuating the context in which they apply to; modeling involves more than just matching a model and a preexisting target. Rather, the modeling activity has a creative part in enabling that one investigates the target, because the target only gets available in precise terms through the applications of the conceptual machinery provided by the model; models and their targets are, in a sense,

co-created. We see the target through the lenses of the model, as it were, and the justification of the model is partly built-in the model, given that the target is framed in theoretical language of the model too:[5]

> in this activity of modelling, the construction of models is intertwined with the construction of new phenomena, theoretical principles and scientific concepts. As a consequence, the justification of a model is partly built into it in the process of modelling, implying that the representational approach, despite its focus on justification, fails to pay enough attention on how models are justified in scientific practice. (Knuuttila and Boon 2011, 311)

In the case of logics, as we have argued, use of a specific modeling of the inferences allowed enables the development of intuitionistic structures (and something similar may be said of classical mathematics, and inconsistent mathematics, with their respective logics). The justification for the use of a given logic is the fact that it is there to begin with, helping us to construct part of the phenomena to be accounted for; the models "both motivate and enable" the construction of the phenomena (Knuutila and Boon 2011, 317). The logics, understood as modeling kinds of inferences, motivate and enable the development of the mathematics associated with them. The same could also be said of classical mathematics, which, in the foundations period, needed to be put in firm basis, by following the standards of the newly developed classical logics. The individuation of the target depends in large measure of the logic used to model the inferences one is interested in. As it happens in science,

> modelling typically involves a theoretical (re)description of the target phenomenon as well as the development of theoretical principles and scientific concepts. The model in the process of its construction functions as an integrating tool as well as a scaffold for further scientific reasoning. In this way the model serves also as a tool of its own development. (Knuuttila and Boon 2011, 316)

In this sense, the development of classical and intuitionistic logics explored the already available knowledge (the controversies on the validity of determined inferences, and the consequences of using only constructive inferences in proofs, for instance), to both be constructed and shape the field being modeled. That is, the model is not only a result of the data we

[5] The idea that models do incorporate 'built-in' justification for suiting their targets comes from Boumans (1999).

put in it, but also it helps us in interpreting and somehow shaping the data, enabling further investigation in terms of the model. This is what happens in classical and intuitionistic mathematics. This is what happens in classical and intuitionistic mathematics. Some of the inferences used in the mathematical practices of the end of the nineteenth century have led to constructions of distinct approaches to the legitimate reasoning in mathematics and, as a result, these advances have enabled the development of distinct mathematical practices itself. The model of inferences is what ends up constraining the development of the field of investigation. The model acts so that it works to delimit the field of application, its phenomena.

Certainly, this only indicates in general lines how a 'logic as models' approach could go, but it does already give us a clear idea that the understanding of context we have suggested can be backed by an approach to models fine-tuned with the current understanding of models in science (being, thus, a naturalistic approach to the methodology of logic too). We suggest that a more pragmatic approach to models in science, which takes seriously the claim that the phenomena is theoretically laden, elaborated in theoretical terms furnished by the model, can have a lot of benefits for logic too. In particular, it can account for the fact that logics are used to generate a plurality of contexts, some of which may be of mathematical interest. Developing further the notion of logic as models in this specific approach is something we leave for some future work.

5. Concluding remarks

We have suggested that Shapiro's approach against a version of generalism in logic seems to face difficulties. We have identified that the major problem seems to be located in an ambiguity as to the role of logic in its relation to the domain or context where the logic is applied. Logic seems to be both used to *characterize* the context, and to be somehow *extracted* from the context. Our proposal to overcome the difficulty consists in locating logic right from the start as an ingredient constituting the domain or context. This makes full sense in the case of mathematical theories, at least as we now conceive of them (and we have discussed only the case of mathematical theories here). Not only does this dissolve the tensions in Shapiro's approach, but also makes room for a more naturalistic approach to the philosophy of logic.

As a by-product, we needed not to enter the discussion of how to grant that a given system is the correct choice for a given context, with disputes typically boiling down to issues of adequacy of systems of logic to the data.

That is, we avoid the kind of discussion concerning whether a logic is right by relating a logic and a preexisting domain, both typically taken as being able to be characterized independently of each other. In our proposal, the correctness of a system is somehow built-in in the very context, and the idea that a context, which is gained so easily, deserves to be studied, depends on pragmatic factors; the decision on which systems are worth of study and development comes from science. Here, of course, Carnapian tolerance is playing a major role.

Furthermore, this approach is nicely suited to the view of logic as models, when 'models' are understood in more naturalistic terms. The view of logic as modeling inferences, and the inferences modeled as delimiting and individuating the field where they apply squares nicely with the localist picture we have advanced. In fact, it boosts the localist proposal advanced here. Advocating a generalist picture, according to this view, would require that one adopts a restrictive position on the domains allowed for an investigation, a restrictive view that is not easy to justify, and that is not justified in the current state of the art of the logic and mathematics as we find it. In this sense, the view advanced here not only helps us advance a more coherent form of localism, but also provides for a clear picture of how new domains come to be proposed, such as paraconsistent mathematics, as we have argued. Certainly, more would still be required to articulate the proposal in all its details, and one may still draw many more important lessons for the epistemology of logic from the use of models in more naturalistic ways, as suggested by the 'models as epistemic tools approach', when connected to the 'logic as models approach', but we leave this issue for another occasion.

Acknowledgments

The paper was prepared with the benefit of a Capes-Humboldt Experienced Researcher Fellowship, held at the Ruhr-Universität Bochum, Germany. The author is partially supported by CNPq (Brazilian National Research Council).

REFERENCES

Boumans, Marcel. 1999. 'Built-in Justification'. In *Models as Mediators. Perspectives on Natural and Social Science,* edited by M. S. Morgan and M. Morrison, 66-96. Cambridge: Cambridge University Press. https://doi.org/10.1017/cbo9780511660108.005

Bridges, Douglas, and Erik Palmgren. 2018. Constructive Mathematics. *The Stanford Encyclopedia of Philosophy,* edited by Edward N. Zalta. Accessed November 1, 2020. https://plato.stanford.edu/archives/sum2018/entries/mathematics-constructive/.

Bueno, Otávio. 2002. 'Can a Paraconsistent Theorist be a Logical Monist?' In *Paraconsistency: The Logical Way to the Inconsistent,* edited by W. Carnielli, M. Coniglio and I. L. D'Ottaviano, 535-552. New York: Marcel Dekker. https://doi.org/10.1201/9780203910139-33

Caret, Colin R. 2021. 'Why Logical Pluralism?' *Synthese* 198: 4947-4968. https://doi.org/10.1007/s11229-019-02132-w.

da Costa, Newton C. A. 1986. 'On Paraconsistent Set Theory'. *Logique et Analyse*, 29 (115): 361-371.

da Costa, Newton. C. A. 1997. *Logiques Classiques et Non Classiques. Essai sur les Fondements de la Logique.* Paris: Masson.

Dicher, Bogdan. forthcoming. 'Requiem for Logical Nihilism; Or, Logical Nihilism Annihilated'. *Synthese.* https://doi.org/10.1007/s11229-019-02510-4.

Haack, Susan. 1978. *Philosophy of Logics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511812866

Hjortland, Ole Thomasson. 2013. 'Logical Pluralism, Meaning Variance, and Verbal Disputes'. *Australasian Journal of Philosophy* 91 (2): 355-373.

Hjortland, Ole Thomason. 2017. 'Anti-exceptionalism about Logic'. *Philosophical studies* 174, 631-658. https://doi.org/10.1007/s11098-016-0701-8

Knuuttila, Tarja, and Mieke Boon. 2011. 'How do Models give us Knowledge? The Case of Carnot's Ideal Heat Engine'. *European Journal for Philosophy of Science* 1: 309-334. https://doi.org/10.1007/s13194-011-0029-3.

Mortensen, Chris. 2017. Inconsistent Mathematics. *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Accessed November 1, 2020. https://plato.stanford.edu/archives/fall2017/entries/mathematics-inconsistent/

Priest, Graham. 1989. 'Classical Logic *Aufgehoben'*. In *Paraconsistent Logic. Essays on the Inconsistent,* edited by G. Priest, R. Routley and J. Norman, chapter 4. Munich: Philosophia Verlag.

Priest, Graham. 2006. *Doubt Truth to be a Liar.* Oxford: Oxford University Press. DOI: 10.1093/0199263280.001.0001

Routley, Richard. 1980. 'The Choice of Logical Foundations: Non-classical Choices and the Ultralogical Choice'. *Studia Logica* 39 (1): 77-98. https://doi.org/10.1007/bf00373098

Shapiro, Stewart. 2014. *Varieties of Logic*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199696529.001.0001

Wyatt, Nicolle; Payette, Gillman. 2021, 'Against Logical Generalism'. *Synthese* 198: 4813-4830. https://doi.org/10.1007/s11229-018-02073-w

BOOK SYMPOSIUM ON THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE: NEW PHILOSOPHICAL AND SCIENTIFIC DEVELOPMENTS BY DEREK BOLTON AND GRANT GILLETT

INTRODUCTION BY GUEST EDITORS

Maria Cristina Amoretti[1] and Elisabetta Lalumera[2]

[1] University of Genoa and PhilHeaD, Research Center for Philosophy of Health and Disease
[2] University of Bologna and PhilHeaD, Research Center for Philosophy of Health and Disease

As is well known, George Engel's seminal paper "The need for a new medical model: A challenge for biomedicine" (1977) argued that medicine should abandon a rigid biomedical model to adopt instead a model that would be able to consider the complex interrelations among the biological, psychological, and socio-environmental determinants of health and disease. Such an interdisciplinary and multidimensional model for addressing the etiology, prevention, prognosis, and clinical treatment of disease is the biopsychosocial (BPS) model.

After more than 40 years, the BPS is taken for granted in some areas of medical research and practice, and at the same time still rejected as vague and ineffective in others. In philosophical quarters the model is equally controversial, as it is welcomed by most anti-dualists, but also targeted by the objections of those who require a mechanistic account of causation, which is still not applicable to psychological-biological and the social-biological relations.

Derek Bolton and Grant Gillett (B&G)'s book starts from acknowledging this partial failure, reviews significant changes that took place in neuroscience, psychology, biology, and healthcare since Engel's proposal, and elaborate a sophisticated defense of the BPS model on philosophical grounds, by providing a new account of the causal relations between the psychological, the biological and the social domain in terms of systems of communication-based regulatory control.

The book is organized into four separate chapters. In the first chapter, B&G present the origin of the BPS model as an alternative to the biomedical model, its long-standing leading role in medicine, healthcare, and health

Correspondence: cristina.amoretti@unige.it; elisabetta.lalumera@unibo.it

educational settings, as well as some recent critiques that have been developed against it, arguing that it is too general, vague, useless, incoherent, and lacking validity. The focus of the second chapter is instead a philosophical argument in favor of a "new biology", which sees biological processes as operating and emerging from information transfer; this argument is in fact needed to dismiss the assumption that only physical causes are "real" causes. The third chapter moves from biology to psychology and is dedicated to discussing the so-called "4-E" model of cognition, which sees cognition as embodied, embedded, enactive, and extended, or ultimately related to agency; within this framework, the "social" component of the BPS model has to do with control and distribution of the resources necessary for biological and psychological life. In the fourth and final chapter B&G argue that the concepts and the boundaries of health and disease are biopsychosocial, utilize the scientific method to identify the causal mechanisms that lead to disease, and identify chronic stress as having a major role in linking psychosocial factors with biological damage. In so doing, they eventually present their renewed BPS model, where physical and mental diseases are brought together, instead of being separate as in the context of the original BPS model.

This book symposium has the aim to further broaden the discussion on the BPS model and its recent reconceptualization through four critical essays.

In the first essay, "From Engel to Enactivism: Contextualizing the Biopsychosocial Model", Awais Aftab and Kristopher Nielsen offer a two-part commentary on B&G's proposal. In the first part, they present a conceptual and historical assessment of the BPS model that is alternative to that offered by B&G, as they take such a model to be less concerned with the ontological possibility and nature of psychosocial causes, and more interested in psychosocial influences. Based on their new assessment, Aftab and Nielsen then question B&G's restricted focus on accounting for biopsychosocial causal interactions. In the second part, B&G's account of mental disorder, which combines the 4E model of cognition with an information-processing paradigm, is compared with a more fleshed out enactivist account of mental disorder that tackles similar conceptual problems of causal interactions but doesn't rely on notions of information-processing.

In the second essay, "Centrifugal and Centripetal Thinking about the Biopsychosocial Model" Kathryn Tabb interprets B&G's reconceptualization of the BPS model as an attempt to increase the conceptual unity of psychiatry. After a brief synopsis of B&G's project and an overview of the main forces currently working against the conceptual unity of psychiatry–forces that have not so much to do with

metaphysical dualism but rather with historic, economic, and sociocultural factors, such as the rise of professional specialization and the related dominance of translational science within psychiatric biomedicine–Tabb argues for psychiatry to acquire a clearly delineated conceptual core. In this respect, she claims, the BPS model should be renewed not only from a metaphysical point of view–as B&G argue–but also, and especially, from an ethical one, as a focus on bioethics could guide choices about *which* causal relationships should be prioritized as research targets in psychiatry.

The third essay, "How to be a Holist Who Rejects the Biopsychosocial Model" by Diane O'Leary, focuses on the BPS model's deeply inconsistent position on dualism, which according to the author may have bad clinical consequences in medicine, too. Very roughly, O'Leary's main point is that it is possible to characterize Engel's driving idea as the acceptance of (phenomenal) consciousness in the context of medical science without retaining the vagueness, platitudeness, and inconsistency of the BPS model itself. This would be possible by embracing metaphysical holism as the willingness to recognize the reality of human experience, and the sense in which that reality forces medicine to address biological, psychological, and social aspects of health. Even if, as O'Leary recognizes, this move will not entirely identify medicine's stance on dualism, it will locate it clearly enough to improve patient care.

In the fourth and final essay, "Causation and Causal Selection in the Biopsychosocial Model of Health and Disease", Hane Htut Maung focuses on some concerns raised by disease causation. To begin, Maung discusses B&G's metaphysical account of biopsychosocial causation, which they see as a preliminary step to defensibly update the BPM model. According to Maung, however, B&G's account is based on claims about the normativity and the semantic content of biological information that are not only metaphysically contentious, but also unnecessary to the scope. On a more general level, moreover, Maung claims that B&G are misdiagnosing the problem, which is not that of providing an adequate account of biopsychosocial causation but that of offering an adequate account of causal selection. He finally considers how the problem of causal selection may be solved to arrive at a more explanatorily valuable and clinically useful version of the BPS model.

The book symposium is closed by Derek Bolton's reply essay, in which he addresses the points raised by the invited commentators.

We wish to thank all the Authors, and especially professor Derek Bolton, for their patience and enthusiasm in this project. We had planned it before the pandemic, not long after the book was published, but many interrelated

causes–as the BPS would have it–postponed its completion for at least one year. We think, however, that a discussion on this important book could not be more timely.

REFERENCES

Bolton, Derek and Gillett, Grant. 2019. *The Biopsychosocial Model of Health and Disease: New philosophical and scientific developments.* London, Palgrave.

Engel, George L. 1977. 'The Need for a New Medical Model: A Challenge for Biomedicine'. *Science,* 196 (4286), 129–136.

EuJAP | Vol. 17 | No. 2 | 2021
UDC: 122:159.97
https://doi.org/10.31820/ejap.17.2.3

FROM ENGEL TO ENACTIVISM: CONTEXTUALIZING THE BIOPSYCHOSOCIAL MODEL

Awais Aftab[1] and Kristopher Nielsen[2]

[1] Case Western Reserve University
[2] Victoria University of Wellington

Original scientific article – Received: 15/01/2021 Accepted: 27/04/2021

ABSTRACT

In this article we offer a two-part commentary on Bolton and Gillett's reconceptualization of Engel's biopsychosocial model. In the first section we present a conceptual and historical assessment of the biopsychosocial model that differs from the analysis by Bolton and Gillett. Specifically, we point out that Engel in his vision of the biopsychosocial model was less concerned with the ontological possibility and nature of psychosocial causes, and more concerned with psychosocial influences in the form of illness interpretation and presentation, sick role, seeking or rejection of care, the doctor-patient therapeutic relationship, and role of personality factors and family relationships in recovery from illness, etc. On the basis of this assessment, we then question Bolton and Gillett's restricted focus on accounting for biopsychosocial causal interactions. The second section compares Bolton and Gillett's account with a recent enactivist account of mental disorder that tackles similar conceptual problems of causal interactions. Bolton and Gillett's utilize elements of the 4E cognition, but they combine these proto-ideas with an information-processing paradigm. Given their explicit endorsement of 4E approaches to mind and cognition, we illustrate some key ways in which a more fleshed out enactive account, particularly one that doesn't rely on notions of information-processing, differs from the account proposed by Bolton and Gillett.

Keywords: *Biopsychosocial model; George Engel; causality; enactivism; 4E cognition*

Correspondence: awaisaftab@gmail.com

"*The Biopsychosocial Model of Health and Disease: New Philosophical and Scientific Developments*" by Derek Bolton and Grant Gillett (2019) is among the most intellectually stimulating books that have been published in the area of philosophy of medicine and philosophy of psychiatry in recent years. It makes notable and substantial contributions to the literature on the biopsychosocial model as well as the nature of causal interactions. It is therefore with pleasure and admiration that we offer this critical commentary.

Our commentary is divided in two sections. In the first section we present a conceptual and historical assessment of the biopsychosocial model (BPSM) that differs from the analysis by Bolton and Gillett (B&G). Specifically, we point out that Engel's BPSM was concerned with much more than the ontological possibility of psychological and social causes. On the basis of this assessment, we then question B&G's restricted focus on accounting for biopsychosocial causal interactions, and in doing so we identify important aspects of debate about the BPSM that we think B&G have overlooked. The second section compares B&G's account with a recent enactivist account of mental disorder that tackles similar conceptual problems. There are aspects of B&G's work that strike us as being somewhat "proto-enactive", although they attempt to combine these ideas with an information-processing paradigm. Given B&G's explicit endorsement of 4E approaches to mind and cognition (Bolton and Gillett 2019, 76), we think it worthwhile to consider the ways in which a fleshed out enactive account differs from the account proposed by B&G.

1. There is More to Engel's BPSM than Causal Interactions

B&G's fundamental focus is on causal interactions in the biopsychosocial realm. They write:

> The conceptual challenge, recognised by Engel and contemporary commentary, is that there are historically deeply entrenched assumptions—physicalism, dualism and reductionism—to the effect that only material, physical and chemical causes are real, while distinctive psychological causes and social causes are impossible or incomprehensible. (Bolton and Gillett 2019, vi)

As such, the majority of their text is focused on developing an account of biopsychosocial causal interactions, the ontological space in which these interactions take place, and how the psychological and social can have genuine causal power within this framework. B&G see their account as a

general model, with the purpose of theorizing biopsychosocial interactions in health and disease. In their words:

> We focus here on the general biopsychosocial model as a core philosophical and scientific theory of health, disease and healthcare, which defines the foundational theoretical constructs—the ontology of the biological, the psychological and the social—and especially the causal relations within and between these domains. (Bolton and Gillett 2019, 19)

B&G are correct that there are historically entrenched assumptions relating to physicalism, dualism, and reductionism that have dominated scientific and medical thinking, and they are also correct that this was recognized by Engel. However, we believe that B&G misdiagnose the negative consequences of these assumptions with which Engel was concerned and which he sought to address in his BPSM. Engel's fundamental concern was not in establishing the reality and existence of psychosocial causes, but rather in the establishing that the psychosocial realm is worthy of scientific exploration and that there is no reason to exclude it from the realm of scientific medical inquiry. Engel was not primarily interested in the alleged impossibility or incomprehensibility of psychological and social causes. We believe this is a fundamental point that has gone by unappreciated not only by B&G, but also in general by commentators following Engel.

That Engel was not primarily concerned with causal interactions is apparent in Engel's seminal papers on BPSM, but becomes even more so when his other writings are considered. In Engel's classic 1977 paper on the subject, Engel is, for a large portion of the article, concerned with the concept of disease and whether our notion of disease should be restricted to biochemical abnormalities. He writes,

> Medicine's crisis stems from the logical inference that since "disease" is defined in terms of somatic parameters, physicians need not be concerned with psychosocial issues which lie outside medicine's responsibility and authority. (Engel 1977)

This statement of medicine's crisis does not indicate a fundamental concern with causal interactions, but rather the nature of our notions of health and disease, and their subsequent implications.

Engel's concerns with the biomedical way of thinking are further expanded on in other articles. In his (1997) article "From Biomedical to Biopsychosocial", Engel sees the aim of the biopsychosocial medicine as being scientific in the human domain:

> Biopsychosocial thinking aims to provide a conceptual framework suitable for developing a scientific approach to what patients have to tell us about their illness experiences (…). Biomedical education's a priori assumption that such patient-derived data and the means of their acquisition are neither teachable, nor subject to systematic study, needs to be examined. (Engel 1997)

Below are some quotations from his (1992) article, "How Much Longer Must Medicine's Science Be Bound by a Seventeenth Century World View?" (Engel 1992)

> In any consideration of a scientific model for medicine that would qualify as a successor to the biomedical model, be it the biopsychosocial or any other, the fundamental issue is whether physicians can in their study and care of patients be scientists and work scientifically in the human domain. Or is medicine's human domain beyond the reach of science and the scientific method, an art, as the biomedical model in effect requires?
>
> Medicine's adherence to a seventeenth century paradigm predicated on the mechanism, reductionism, determinism, and dualism of Newton and Descartes automatically excludes what is distinctively human from the realm of science and the scientific.
>
> Biomedicine's rejection of dialogue as a genuinely scientific means of data collection is evident in the neglect of instruction and supervision in interviewing, not to mention in clinical data collection altogether, and in the preference for the case presentation as a method of clinical teaching, one in which students may display their ability to organize and discuss findings, but not reveal the methods and skills whereby they had come by the data in the first place, least of all their interpersonal engagement with the patient.

This is recognized, to an extent, even by B&G, because they begin chapter 1 by listing what Engel identified as limitations of the biomedical model, that it fails to take into account the following:

> the person who has the illness, the person's experience of, account of and attitude towards the illness; whether the person or others in fact regard the condition as an illness; care of the patient as a person; for some conditions such as schizophrenia

> and diabetes, the effect of conditions of living on onset, presentation and course; and finally, the healthcare system itself also cannot be conceptualised solely in biomedical terms but rather involves social factors such as professionalization. (Bolton and Gillett 2019, 2)

Notably, concerns about the *causal reality* of psychosocial factors do not appear on this list by Engel, because such concerns are prominently missing from Engel's seminal writings. Given Engel's strong interest in the various dimensions of the illness experience and utilizing the clinical interview as an instrument of scientific inquiry, it is quite possible that Engel would have been dismayed to see interpretations of BPSM as having to do primarily with causal interactions.

It needs to be stated that the responsibility for this misunderstanding of Engel's thesis doesn't lie with B&G. Such a characterization of BPSM as being concerned primarily with *causes* is widespread, even among the most ardent champions of BPSM. Consider, for instance, Dr Ronald Pies, author of *Clinical Manual of Psychiatric Diagnosis and Treatment: A Biopsychosocial Approach* (Pies 1994), who wrote in *Psychiatric Times* in 2020: The biopsychosocial paradigm

> asserts that most (but not necessarily all) serious mental disorders are best understood as having a variety of causes and risk factors—including but not necessarily limited to biological, psychological and social components. (Pies 2020)

While such a formulation is not strictly erroneous, it is a more restrictive understanding of Engel's vision (Aftab 2020). The matters that preoccupy Engel are more to do with psychosocial influences in the form of illness interpretation and presentation, sick role, seeking or rejection of care, the doctor-patient therapeutic relationship, and role of personality factors and family relationships in recovery from illness, etc. Engel was seeking a framework that would bring the psychosocial and phenomenological dimensions of illness within the realm of medical and scientific inquiry. Causes and risk factors are included in it, surely, but they are not particularly privileged by Engel.

Why then has our popular understanding of BPSM been so focused on causal risk factors and causal interactions? This appears to be a consequence of the manner in which BPSM has been operationalized and taught to medical trainees. The operationalization has taken the form of a biopsychosocial formulation. This formulation is illustrated as a table in which there are three columns of "biological", "psychological" and

"social", and four rows of predisposing factors, precipitating factors, perpetuating factors, and protective factors (see Huda 2020 for an example of such a formulation). This organization urges the trainees to take into account all the various causal factors by filling in all the boxes. Furthermore, such a formulation is intended to assist in the development of a treatment plan, with the understanding that the treatment should be aimed at all the modifiable causal factors identified.

The biopsychosocial formulation, while a useful educational and clinical tool, creates a number of conceptual and philosophical problems (Waterman 2006). First of all, it encourages the reification of "biological", "psychological" and "social" as separate and distinct ontological domains. Such a reification is illusory, since there are good reasons to think that the biological, the psychological, and the social as levels of explanation are best understood as heuristic idealizations that are helpful in making certain sorts of distinctions of interest to us, but do not reflect deep ontological features of the world (see Eronen 2021 for a defense of this view). Secondly, causal factors identified have to be cleanly sorted into one or the other column, often in an arbitrary or artificial manner (e.g., is "pain" a biological or a psychological factor?). Thirdly, while all the risk factors are categorized, no weight is assigned regarding their respective causal roles, giving the (false) impression that they "are all, more or less equally, relevant". Fourthly, since a combination of bio-psycho-social factors will almost always be present, a clinician may feel justified in offering any sort of treatment that is *perceived* to address those factors, regardless of whether that treatment is backed by scientific evidence or is recommended by guidelines. Fifthly, creating a static array of causal risk factors further enhances the mystery of how these causal factors interact dynamically in the real world.

It is in the face of such an understanding of BPSM that Paul McHugh and Philip Slavney (1998) argue that the model is amorphous and vague, offering little meaningful guidance for clinical and research work. They see BPSM as analogous to a list of ingredients rather than a recipe, providing no instructions on how these ingredients are to be effectively mixed together in the process of cooking.

It is also important to understand the ideological function that BPSM has served in psychiatry. BPSM was utilized as a means of bridging the rift between the various factions within psychiatry with biological, psychological, and social orientations (Ghaemi 2010). It did so by a sort of Dodo bird verdict that all approaches are legitimate, and none shall be excluded, "everyone has won, and all must have prizes". It is this rhetorical function of BPSM that leads Ghaemi (2010) to contend that in

contemporary practice BPSM has led the clinicians into a state of lazy eclecticism.

While B&G allude to some of this, and recognize that the attitude of uncritical eclecticism is not present in Engel's original writings, they fall short in two ways: i) they don't recognize, at least explicitly, that a central preoccupation with causal interactions is also not present in Engel's writings, and ii) they don't seem to demonstrate an adequate appreciation that many criticisms of BPSM are directed at the manner in which BPSM has been operationalized and implemented. Given this targeting, such criticisms will stand as long as the practical implementation of BPSM remains the same.

While B&G highlight the criticisms of BPSM by Ghaemi and Kendler, they don't seem to make much effort at engaging with the conceptual alternatives offered by these authors. Both Ghaemi and Kendler endorse versions of "pluralism" as replacements for BPSM, Jasperian methodological pluralism in the case of Ghaemi (2010), and explanatory integrative pluralism in the case of Kendler (2005). The basic viewpoint of such pluralisms is that multiple independent methods and explanations (at multiple levels/scales) are necessary to understand and treat mental illnesses. The strengths and limits of each method or explanation need to be recognized, and that method/explanation should be utilized which is best suited for the specific circumstances based on pragmatic constraints, relevant epistemic values, and empirical evidence.

There is somewhat of a parallel here to B&G's assertion that the content of the BPSM is in the *specifics*. It can be argued that saying that the content of BPSM lies in the scientific and clinical specifics is not that much different from saying that our understanding of specific conditions and disorders should be guided by the best available scientific explanations for those disorders, explanations which will almost always include psychosocial variables in addition to biological variables, either as contributing to etiology, presentation, course, or treatment considerations. The value that BPSM offers in this regard is basically as a reminder: do not restrict your notions of scientific inquiry to exclude the human and the psychosocial realm. Aside from serving as a reminder, it does not seem to offer anything above and beyond what we would expect a good scientific explanation to offer. In other words, a good scientific explanation of a complex, multifactorial medical condition such as diabetes or depression will invariably be one that includes biological, psychological, and social variables, but that is not because the good scientific explanation will be derived from BPSM.

In a similar vein, the value of BPSM in clinical practice and medical education is that of a *reminder* not to ignore psychosocial variables. Such a reminder is necessary because of medicine's long-trenched history of focusing on the biological to the exclusion of the psychosocial. As noted by Kendler:

> [BPSM] is used widely in family medicine and is a great teaching tool, reminding the residents to consider the psychological and social influences on their cases and not just focusing on the pathophysiology. (Kendler 2010)

A philosophical account of bio-psycho-social causal interactions doesn't quite serve the same purpose. This also indicates that when it comes to BPSM as it currently exists, calling it a "model" is beyond charitable (McLaren 1998). It is more of an attitude, a mantra, a meditation, a nudge, an aide-memoire, rather than anything as elaborate as a "model", and assuming that it is indeed a model creates all sorts of conceptual problems. B&G's philosophical account of biopsychosocial causal interactions is a worthwhile philosophical inquiry, but in light of Engel's original writings, there is no good reason that BPSM should concern itself solely with causal interactions, to the exclusion of issues that were of concern to Engel: the human domain with all its quirks and colors. Even if a successful account of biopsychosocial interactions were to be provided, it does little to address the conceptual and scientific issues in contemporary practice of, in the words of Kendler, "how to integrate the diverse etiologic factors that contribute to psychiatric illness and how to conceptualize rigorously multidimensional approaches to treatment" (Kendler 2010). Establishing the psychological and the social as ontologically and causally real doesn't help us with the question of how to best integrate the etiological factors in the form of a coherent explanation and how this should inform multidimensional approaches to treatment.

In summary of section 1:

- *An interpretation of BPSM with a central emphasis to causal interactions is at odds with Engel's vision of BPSM which was focused more on bringing the human domain into the scientific realm, establishing clinical interview as a scientific instrument, taking illness experience seriously as scientific data, and adopting a non-reductionistic view of disease and health.*

- *Many popular criticisms of BPSM are targeted at how BPSM has been operationalized and implemented for the purposes*

of clinical education, and the way the rhetoric of BPSM has been used for ideological purposes. Reinterpreting BPSM as a philosophical account of biopsychosocial causal interactions will not, by itself, address these concerns.

- *The assertion that the content of the BPSM is in the specifics does not seem to offer anything above and beyond what we would expect a good scientific explanation to offer. In other words, a good scientific explanation of a complex, multifactorial medical condition such as diabetes or depression will invariably be one that includes biological, psychological, and social variables, but that is not because the good scientific explanation will be derived from BPSM.*

- *Given the historical dominance of the reductionistic scientific worldview, BPSM appears to serve as a reminder to avoid the reductionistic trappings of the biomedical mindset; its clinical and educational value appears to be as a mantra, a nudge, an aide-memoire, rather than anything as elaborate as a "model", and assuming that it is indeed a model creates all sorts of conceptual problems.*

- *Establishing the psychological and the social as ontologically and causally real doesn't help us with the question of how to best integrate the etiological factors in the form of a coherent explanation and how this should inform multidimensional approaches to treatment.*

- *B&G do not seem to pay attention to the alternatives to BPSM that have emerged in the last 2 decades in the philosophical literature, such as various forms of explanatory and methodological pluralisms.*

2. Comparison with an Embodied Enactive View

As conceptual pluralists, we see value in there being a variety of ways to view something as complex as health and well-being. However, these different views must be allowed to 'bounce off' each other—to be compared in terms of strengths and weaknesses and refined in response. It is through diversity *and* dialogue that better frameworks will emerge. In this section we compare B&G's BPSM to one such developing alternative, the embodied, embedded, and enactive view of psychopathology (3EP)

(Nielsen 2020, 2021; Nielsen and Ward 2018, 2020). As we have mentioned earlier, B&G cite the 4E framework as inspiration for their own view of embodied agency, but there are substantial differences between their model and models of health and disease that have emerged from, identify with, and operate within the 4E tradition.

Very briefly, 3EP is an approach to conceptualizing mental disorder grounded in a view of human functioning as embodied (fully material, and constituted by not just the brain, but the brain-body system), embedded (richly and bi-directionally connected to the world around us), and enactive (meaning is not out there in the world, nor is it 'constructed' by us, but rather concerns the very real relation between the state of the world and our purpose to try to keep living). While being a 'biological' position that acknowledges the importance of physiological processes for understanding behavior, 3EP places equal value on personal meaning and interpersonal scales of explanation. In this way it is a non-reductionistic position, yet does not ignore the importance of the body and its biological constitution. 3EP thus sees all the various scales of explanation relevant to understanding human behavior as different perspectival aspects of the same dynamic whole – an organism standing in relation to its environment (both physical and socio-cultural). On this view mental disorders appear as patterns existing across brain, body, and environment, keeping people stuck in patterns of behavior that are working against their own adaption and self-maintenance. To conserve space this summary has been extremely brief. For fuller accounts see: Nielsen (2020, 2021), Nielsen and Ward (2018, 2020). For a complimentary perspective on mental disorder referred to as Enactive Psychiatry see: de Haan (2020a, 2020b, 2020c).

While the BPS is a general framework of health and 3EP is a developing conceptual perspective specifically focused on mental disorder, both positions overlap in important ways. Both positions seek to move beyond purely biomedical understandings and recognize the legitimacy of socio-cultural and environmental impacts on health. Further, both do so by claiming to place biological, psychological, social, and environmental factors into a single ontological space, thus accounting for increasingly recognized interactions between these 'domains'. Both positions engage with notions of formal/organizational causality as seen through their shared talk of 'systems' and 'dynamics'. Finally, both positions seem to see such organizational causality as a way to account for the emergence of apparent purposes/teleology, against which they can meaningfully speak of function/dysfunction. There are however, important differences in how these tasks are achieved. Here we will explore two of these differences, and use the discussion to highlight areas where the current construal of Bolton and Gillett's BPS leaves us wanting to know more.

2.1. The Role of 'Information'

Following Engel, Bolton and Gillett's BPS framework views the world in terms of relatively distinct (but not ontologically separate) domains of the biological, the psychological, and the socio-political. This then presents them with somewhat of a 're-stitching' problem, and they subsequently account for relationships between these domains using the key notions of *information transfer* and *regulatory control*. At the risk of over-summarizing this view: Biological processes receive information/instructions from DNA and, through following these instructions, regulate their own physico-chemical constitution and immediate environments. Psychological processes meanwhile (embodied in the nervous system) receive and integrate information about the state of the self and the world via sensory input, and attempt to regulate the world and self in a way that meets the organism's needs through embodied agency. Finally, socio-political processes (embodied in the actions of the collective) involve the perception and recognition of others (a complex form of information transfer), and the regulatory control of resources needed by individuals.

An important question at this point however is 'what exactly is information?'. The notion of information in biological systems has generated considerable philosophical debate, and these debates are of great relevance to B&G given the central role information plays in their account. Godfrey-Smith and Sterelny's (2007) entry on "Biological Information" in *Stanford Encyclopedia of Philosophy* is a great resource for this purpose, and we'll summarize some pertinent remarks here. An uncontroversial and minimal notion of information is that of Shannon information, according to which any variable may be said to 'contain/carry/be' information about a source if it correlates with the state of that source. On this account information is said to be present in the variable in that the variable can be used to predict the state of the distal source. There is no greater commitment in Shannon information that there is any biological system designed/intended to produce that signal or to use it once produced. Biologists, however, often appear to use a notion of information that is richer than Shannon information and much more controversial, i.e. information with semantic and intentional content. Godfrey-Smith and Sterelny (2016) present readers with three options with regards to the concept of semantic information in biology:

1. Semantic information is useful as an analogy, as a metaphor, but not intended to be literally true.

2. Semantic information literally exists in biological systems, in which case the task of the philosopher is to explain how

semantic information can arise and exist in non-intelligent systems.

3. Shannon information is sufficient for biological systems and no richer concept is needed.

We don't intend to settle this debate here or defend a particular approach, but we want to point out that the philosophical validity of any particular view is far from obvious. It would appear that B&G would adopt the second view, that semantic information literally exists, but it is unclear how they would defend it. B&G do, however, demonstrate clear awareness of the contextuality of information. For example, when discussing genetics/DNA they stress that

> genes code for particular proteins (…) [where] 'code for' means: in normal circumstances, in the normal cellular environment, in a complex series of interlocking steps, such-and-such DNA sequence produces such-and-such protein. (Bolton and Gillett 2019, 54)

In making such specifications they acknowledge awareness that information is always contextual—e.g., language is gibberish to those of a completely different social-cultural context. Ultimately information is merely a flow of change within a system, change that is then used by the system in some way. This would suggest that their view is also compatible with understanding semantic information as analogy, an epistemological tool utilized by observers—a way that we can make (our own) sense of the system/s understudy. As such, information-processing is a model or metaphor, representing one possible way to understand a system. Either way, there is little philosophical clarity on this point.

No such information processing metaphor is employed under the 3EP view. Under 3EP there is no tripartite structure to the ontology. Instead, the brain, body, and environment are considered to all be constituted from material substance, and to form a complex dynamical system existing across different scales of time and space—i.e., the so-called 'brain-body-environment system'. Rather than traditional levels of ontology such as the genetic, cellular, organistic, organismic, behavioral, or social, 3EP recognizes such divisions as simply referring to increasing constitutional complexity across increasing scales of time and space, with the emergence of some organizationally closed systems along the way (Di Paolo et al. 2018; Maiese 2016; Thompson 2007; Varela et al. 2017; Potochnik 2010). Because of this there is no mysterious interaction between domains or levels to be explained by information exchange. Thus, instead of the

language of 'information' and 'regulatory control' seen in the BPS, 3EP utilizes the language of *organizational* or *circular causality* (Fuchs 2017), speaking of concepts such as *emergence*, *constraint*, and *constitution*, when navigating multi-scale interactions.

A question that may arise at this point is, what then is the psychological in such a materialist (but dynamical) worldview? In short, under the enactive approach the biological and psychological are seen as continuous. The psychological is something that is enacted through the organization and action of the biological organism (Thompson 2007). To put it another way, the enactive approach avoids substance dualism by holding 'the mind' to be a verb, not a noun. This relates to a key concept of the enactive approach known as the 'deep continuity thesis', which we will return to in the next section. On this view the organizational structures of life *are* the structures of mind and the psychological is therefore thoroughly embodied.

As one way of attempting to understand the dynamic constitution of a human being standing in their environment, the model of information processing may well be a helpful one. In essence it represents somewhat of a cognitive/epistemological short-cut via metaphor to communications equipment or computers. However, B&G reference the idea of an embodied, embedded, and enactive mind as inspiration for their framework, and these ideas apparently play a core role in their concept of embodied agency (Bolton and Gillett 2019, 76). Given that these schools of thought commonly avoid talk of information, and arguably successfully navigate similar conceptual issues to the BPS without reliance on an information-processing metaphor, the necessity of B&G's reliance on an information processing approach is not entirely clear.

2.2. The Emergence of Normativity/Functionality

One of the biggest challenges for naturalist approaches to conceptualizing health is that health is a fundamentally normative idea, and the natural is traditionally seen in opposition to the normative. In order to say that some state of the world is naturally preferable to another (e.g., not having cancer vs. having cancer) we need to be able to traverse the 'normative gap' between what is (i.e., the factual state of a person) and what we are claiming ought to be (i.e., a state of health). B&G's biopsychosocial framework claims to have crossed this divide. For example, they claim that "(…) the theory is fundamentally normative (…)" (Bolton and Gillett 2019, 35). However, as far as we can tell they do not directly and explicitly address how they see themselves as having crossed it. Within the biological domain they appear to attempt to do so using the notion of information and error. As they move into the psychological and socio-political domains

they appear to shift to a reliance on a systems-based notion of functionality and preservation of the system. In this section we compare B&G's approach to the 3EP approach which is more thoroughly systems-based and currently more specified. We argue that this systems-based understanding is preferable, and that the BPS could be improved by explicitly and more thoroughly assuming such a systems-based approach.

In chapter 2 while discussing the biological domain, B&G state that

> (…) regulation and control mechanisms keep things going *right rather than wrong*. Such normativity is not apparent in the energy equations of physics and chemistry, which always apply and never fail. It arises in biology for the first time, marking a fundamental departure of biology from physical and chemical processes alone. (Bolton and Gillett 2019, 50)

They also seem to imply that this normativity has to do with information and how it can contain errors or be misread

> (…) the information-processing paradigm in biology secures the fundamental point that the functional end of a system (…) is (…) already present in the system prior to production, as instructions and a mechanism for the production. (Bolton and Gillett 2019, 54)

It is therefore through the fact that we can see 'instructions' in biology/DNA that B&G claim we can first see normativity arising.

However, B&G also reference a different source of normativity, that of the wider functioning of the system. They state that "(…) normativity also applies at the level of the whole organism in interaction with the environment: interaction is *adaptive* insofar as it promotes continuity and functioning and is otherwise *maladaptive*" (Bolton and Gillett 2019, 51). As B&G shift to discussing the psychological and social domains in chapter 3, and the wider notions of health and disease in chapter 4, they appear to speak less about information and error as a normative basis, and more about perpetuation of the system as a basis for defining functionality. For example, in chapter 4, when they come closest to directly addressing the source of normativity within the BPS, they are clear that the logic of attributing disease is 'top-down'. They state that "[i]t is poor outcomes at the level of the whole that ultimately drives attribution of dysfunctionality downwards to the parts that serve the whole" (Bolton and Gillett 2019, 111).

The 3EP perspective has a strength in that it directly addresses this normative gap. Nielsen and Ward (2020) explore how the enactive concepts of self-maintenance and adaption, grounded in the organizational structures of life, lay the groundwork for a view of mental disorder that is both natural *and* normative. In doing so, they also draw on the work of non-enactive authors that have developed consilient arguments for the natural emergence of normativity such as Okrent (2017) and Christensen (2012). They demonstrate how the deep continuity thesis at the heart of enactivism is itself an account of natural normativity:

> Under the deep continuity thesis, all life shares an embodied "concern" (i.e., a self-perpetuating structure) for the continuation of the self (…) in the face of changing and precarious environmental conditions (…). Insofar as an organism *should* act to maintain its own life, there are states, actions, and processes that the organism *should* be in or perform. (Nielsen and Ward 2020, 8)

From these roots, Nielsen and Ward show how a view emerges where mental disorder can be seen as a pattern of behavior (including cognition and affect), enacted by an organism, that pushes significantly counter to its own self-maintenance and adaption in context.

Such a perspective aligns well with a view where organisms are understood as systems that maintain a non-equilibrium steady state, temporarily pushing back against the 2nd law of thermodynamics. Coming at the same idea from this explicitly systemic view, what is functional is what manages to serve the survival of the organism at a non-equilibrium steady state within a fluctuating environment. A similar systemic notion of functionality appears to be inherent (and potentially extended) in recent perspectives such as active/enactive inference (Ramstead et al. 2020), or the social ecological model of mental functioning (Chapman 2021). As mentioned, such a view is alluded to by B&G but is currently somewhat underspecified. Given our concerns about the role of information expressed in the previous section, we suspect this systemic approach holds much greater potential than attempting to ground normativity in the idea of information and error.

2.3. Summary

In summary of section 2:

- *B&G explicitly reference ideas of embodiment, embedment, and enactivism, and their work shares some overlap in intention with*

a 3EP approach. Their work seems somewhat 'proto-enactive' in that these ideas are referenced but do not seem to permeate their approach.

- *B&G's notion of 'information' is currently underspecified and potentially in tension with their supposed grounding in ideas of embodiment and enactivism.*

- *B&G claim to have crossed the 'normative gap', a challenge for any naturalist account of health and disease, but how they do so is unclear.*

- *At times, B&G seem to reference a systems-based/organizational notion of natural normativity. Such an approach has potential, but is significantly underspecified in their current account. Such an approach is more fully explored by Nielsen and Ward (2020).*

Acknowledgments

The authors would like to thank Hane Maung for his comments on a draft of this article

REFERENCES

Aftab, Awais. 2020. The Nine Lives of Biopsychosocial Framework. *Psychiatric Times*. Accessed July 24, 2021. https://www.psychiatrictimes.com/view/nine-lives-biopsychosocial-framework

Bolton, Derek and Grant Gillett. 2019. *The Biopsychosocial Model of Health and Disease: New Philosophical and Scientific Developments*. Cham, Switzerland: Palgrave Pivot.

Chapman, Robert. 2021. 'Neurodiversity and the Social Ecology of Mental Functioning'. *Perspectives on Psychological Science:* 1745691620959833. https://doi.org/10.1177/1745691620959833.

Christensen, Wayne. 2012. 'Natural Sources of Normativity'. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43 (1): 104–112. https://doi.org/10.1016/j.shpsc.2011.05.009.

de Haan, Sanneke. 2020a. 'An Enactive Approach to Psychiatry'. *Philosophy, Psychiatry and Psychology* 27 (1), 3–25. https://doi.org/10.1353/ppp.2020.0001.

de Haan, Sanneke. 2020b. 'Bio-psycho-social Interaction: An Enactive Perspective'. *International Review of Psychiatry*, 1–7. https://doi.org/10.1080/09540261.2020.1830753.

de Haan, Sanneke. 2020c. *Enactive Psychiatry*. Cambridge, UK: Cambridge University Press.

Di Paolo, Ezequiel A., Elena Clare Cuffari, and Hanne De Jaegher. 2018. *Linguistic Bodies: The Continuity between Life and Language*. Cambridge, MA: The MIT Press.

Engel, George L. 1977. 'The Need for a New Medical Model: A Challenge for Biomedicine'. *Science* 196 (4286), 129-136. https://doi.org/10.1126/science.847460.

Engel, George L. 1992. 'How Much Longer Must Medicine's Science Be Bound by a Seventeenth Century World View?'. *Psychotherapy and Psychosomatics 57* (1-2), 3-16. https://doi.org/10.1159/000288568.

Engel, George L. 1997. 'From Biomedical to Biopsychosocial: Being Scientific in the Human Domain'. *Psychosomatics* 38 (6), 521-528. https://doi.org/10.1016/S0033-3182(97)71396-3.

Eronen, Markus I. 2021. 'The Levels Problem in Psychopathology'. *Psychological Medicine* 51 (6), 927-933. https://doi.org/10.1017/S0033291719002514.

Fuchs, Thomas. 2017. *Ecology of the Brain: The Phenomenology and Biology of the Embodied Mind*. Oxford, UK: Oxford University Press.

Ghaemi, S. Nassir. 2010. *The Rise and Fall of the Biopsychosocial Model: Reconciling Art and Science in Psychiatry*. Baltimore, MD: JHU Press.

Godfrey-Smith, Peter and Kim Sterelny. 2016. *Biological Information*. The Stanford Encyclopedia of Philosophy. Edited by Edward N. Zalta. Accessed July 24, 2021. https://plato.stanford.edu/archives/sum2016/entries/information-biological/.

Huda, Ahmed Samei. 2020. 'The Medical Model and Its Application in Mental Health'. *International Review of Psychiatry* (2020): 1-8. https://doi.org/10.1080/09540261.2020.1845125.

Kendler, Kenneth S. 2005. 'Toward a Philosophical Structure for Psychiatry'. *American Journal of Psychiatry* 162 (3): 433-440. https://doi.org/10.1176/appi.ajp.162.3.433.

Kendler, Kenneth S. 2010. 'The Rise and Fall of the Biopsychosocial Model: Reconciling Art and Science in Psychiatry (Book Forum)'. *American Journal of Psychiatry* 167 (8): 999. https://doi.org/10.1176/appi.ajp.2010.10020268.

Maiese, Michelle. 2016. *Embodied Selves and Divided Minds*. Oxford: Oxford University Press.

McHugh, Paul R., and Phillip R. Slavney. 1998. *The Perspectives of Psychiatry*. Baltimore, MD: JHU Press.

McLaren, Niall. 1998. 'A Critical Review of The Biopsychosocial Model'. *Australian & New Zealand Journal of Psychiatry* 32 (1), 86-92. https://doi.org/10.3109/00048679809062712.

Nielsen, Kristopher. 2020. *What is Mental Disorder? Developing an Embodied, Embedded, and Enactive Psychopathology* [PhD thesis, Victoria University of Wellington]. Accessed July 24, 2021. http://hdl.handle.net/10063/8957.

Nielsen, Kristopher. 2021. 'Comparing Two Enactive Perspectives on Mental Disorder'. *Philosophy, Psychiatry and Psychology*, 28 (3), 175-185. https://doi.org/10.1353/ppp.2021.0028.

Nielsen, Kristopher, and Tony Ward. 2018. 'Towards a New Conceptual Framework for Psychopathology: Embodiment, Enactivism and Embedment'. *Theory & Psychology*, 8 (6): 800–822. https://doi.org/10.1177/0959354318808394.

Nielsen, Kristopher, and Tony Ward. 2020. Mental Disorder as both Natural and Normative: Developing the Normative Dimension of the 3E Conceptual Framework for Psychopathology. *Journal of Theoretical and Philosophical Psychology* 40 (2): 107–123. https://doi.org/10.1037/teo0000118.

Okrent, Mark. 2017. *Nature and Normativity: Biology, Teleology, and Meaning*. New York: Routledge.

Pies, Ronald. 2020. Can We Salvage the Biopsychosocial Model? *Psychiatric Times*. Accessed July 24, 2021. https://www.psychiatrictimes.com/view/can-we-salvage-biopsychosocial-model.

Potochnik, Angela. 2010. Levels of Explanation Reconceived. *Philosophy of Science* 77 (1): 59–72. https://doi.org/10.1086/650208.

Ramstead, Maxwell J. D., Michael D. Kirchhoff, and Karl J. Friston. 2020. 'A Tale of Two Densities: Active Inference Is Enactive Inference'. *Adaptive Behavior* 28 (4): 225–239. https://doi.org/10.1177/1059712319862774.

Thompson, Evan. 2007. *Mind in Life: Biology, Phenomenology, and the Sciences of Mind*. Cambridge, MA: Harvard University Press.

Varela, Francisco J., Evan Thompson, and Eleanor Rosch. 2017. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: The MIT Press.

Waterman, G. Scott. 2006. 'Does the Biopsychosocial Model Help or Hinder Our Efforts to Understand and Teach Psychiatry?'. *Psychiatric Times* 23 (14): 12-13.

EuJAP | Vol. 17 | No. 2 | 2021
UDC: 179:616-89
https://doi.org/10.31820/ejap.17.2.4

CENTRIFUGAL AND CENTRIPETAL THINKING ABOUT THE BIOPSYCHOSOCIAL MODEL IN PSYCHIATRY

Kathryn Tabb[1]

[1]Philosophy Program, Bard College

Original scientific article—Received: 25/07/2021 Accepted: 21/09/2021

ABSTRACT

The biopsychosocial model, which was deeply influential on psychiatry following its introduction by George L. Engel in 1977, has recently made a comeback. Derek Bolton and Grant Gillett have argued that Engel's original formulation offered a promising general framework for thinking about health and disease, but that this promise requires new empirical and philosophical tools in order to be realized. In particular, Bolton and Gillett offer an original analysis of the ontological relations between Engel's biological, social, and psychological levels of analysis. I argue that Bolton and Gillett's updated model, while providing an intriguing new metaphysical framework for medicine, cannot resolve some of the most vexing problems facing psychiatry, which have to do with how to prioritize different sorts of research. These problems are fundamentally ethical, rather than ontological. Without the right prudential motivation, in other words, the unification of psychiatry under a single conceptual framework seems doubtful, no matter how compelling the model. An updated biopsychosocial model should include explicit normative commitments about the aims of medicine that can give guidance about the sorts of causal connections to be prioritized as research and clinical targets.

Keywords: *Biopsychosocial model; precision medicine, medical ethics; philosophy of psychiatry*

Correspondence: ktabb@bard.edu

1. Introduction

Writing on the tortured status of psychiatric classification, Scott Lilienfeld (2014) characterized the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) as buffeted about by conflicting centrifugal and centripetal forces. Often psychiatric nosology is envisioned in awkward suspension between the twin stars of Snow's two cultures. Lilienfeld's metaphor has it instead shifting unstably amidst the ongoing negotiations of a range of subtler powers. For my purposes I will borrow the metaphor not—or not just—in order to reflect on the shaky orbit of the DSM around the nebula of scientific validity, but rather in order to say something about the shifting conceptual structure of the discipline of psychiatry as a whole. The centripetal forces I am interested in are those compressing the field of psychiatry into some sort of conceptual unity. The centrifugal ones are those pulling it apart, as some bits spin off into the basic and applied sciences, and others move farther into humanistic spaces like psychotherapeutics, recovery movements, and social welfare projects. Going back to Jaspers, a worry that psychiatry has two distinct projects that are increasingly uneasy together—one that values explanation, and one that values understanding—has driven scholars and clinicians to offer up various pleas for centripetalism, the calling back to order of an undisciplined discipline. I am thinking of titles like David Brendel's *Healing Psychiatry*, or Tanya Luhrmann's *Of Two Minds*. Many of these centripetal pleas attribute this historic split to the broader Cartesian severing of the ontological into the physical and the mental, which, they claim, has destabilized psychiatry, balanced as it is on the point where the two meet.

Perhaps most notable among such attempts has been the biopsychosocial model, introduced by George L. Engel in 1977. If it still functions as a model for psychiatry—rather than as something more like a zeitgeist—it does so in an optative mood; not so much supplying a rigorous descriptive or prescriptive representation of contemporary medicine as offering a cultivated and relatively benign rebuke to the way things are. In their monograph *The Biopsychosocial Model of Health and Disease: New Philosophical and Scientific Developments*, Derek Bolton and Grant Gillett aim to realize some of the model's original transformative potential, not only for psychiatry but for medicine writ large. Through integrating not only our best contemporary theories of each level of analysis—the biological, the psychological, and the social—but also our best theories of their concomitance, the authors aim to save the model from the aggregated charges of imprecision, disappointing scientific validity, and philosophical incoherence that have built up over decades (Bolton and Gillett 2019, v).

I am sympathetic to the anxieties about centrifugalism that have increasingly animated philosophers of psychiatry. I am also galvanized by Bolton and Gillett's case for drawing our attention back to the biopsychosocial model's original promise, on the grounds that we now have the scientific and philosophical tools to make it work better. In response, I want to offer some reasons for thinking that the centripetal force that Bolton and Gillett posit—a fundamentally *metaphysical* force—may not sufficiently address some of prevalent worries about psychiatry's current predicaments (I think it is the case that these worries are also applicable to much of contemporary medicine, such that the shortcomings I see in their model would apply in other contexts as well, but here I limit my discussion to psychiatry). In particular, I will argue that *ethical* arguments for centripetalism are necessary alongside metaphysical ones, and that therefore, if the biopsychosocial model is to be resuscitated, it should be resuscitated in a manner that gives ethical forces primacy. I will not, for the most part, engage with the details of Bolton and Gillett's argument, which I think are rich and exciting, and which I expect will prompt a great deal of interest from philosophers working at the interstices of explanation, causation, and philosophy of mind. Little I say here conflicts with the nuts and bolts of their new model, but I do want to shift the center of gravity a bit.

In the following section I give a brief synopsis of Bolton and Gillett's project, a true challenge given the density and richness of their slim book. In Section 3 I will review what I see as the main forces working against conceptual unity in psychiatry, and review the strongest grounds, as I understand them, for worries that the discipline increasingly lacks a clearly delineated conceptual core. I will argue that this is less about dualism—indeed, less about philosophy!—than about historic, economic, and sociocultural factors which have motivated different practitioners to adopt different competing conceptual schemata. In particular I will highlight the dramatic rise of professional specialization within the field of psychiatry during the twentieth century, and the related dominance of translational science over clinical science within psychiatric biomedicine. In Section 4 I will discuss how a focus on bioethics could complement the new biopsychosocial model by guiding choices about *which* causal relationships should be prioritized as research targets in psychiatry. Finally, I will conclude with some reflections on what it might look like to integrate ethical principles into the new biopsychosocial model such that they, too, would act as a centripetal force.

2. Ontological Centrifugalism, Ontological Centripetalism

Bolton and Gillett's case must start with persuasive evidence that the biopsychosocial model is worth restoring. Their project responds to critics like Nassir Ghaemi, who frames his *Rise and Fall of the Biopsychosocial Model* around the arresting claim that the model in its original psychiatric context "rose from the ashes of psychoanalysis and is dying on the shoals of neurobiology" (2010, ix). Engel's intended intervention indeed arose from the contingencies of its historical moment—by the nineteen seventies the conflagration, or sea change, from the old psychoanalytic paradigm that had shaped the first edition of the DSM in 1952 to the operationalism that guiding the production of the DSM-III (1980) was well underway. The optimism over psychiatry's status as a science, which led to the emphasis on objective observation in the manual's third edition, was due in part to recent discoveries of powerful new psychotropic drugs. While these advances were not, actually, born of new insights into the causal mechanisms underlying mental illness, they gave reason to hope that scientific breakthroughs would be forthcoming. Engel's biopsychosocial model was intended to counter the rising enthusiasm for defining disease exclusively in terms of "somatic parameters", not only in psychiatry but in medicine as a whole (Engel 1992, 317). At a time when many psychiatrists were desperate to justify psychiatry as a legitimate medical science even as the care of the mentally ill was increasingly handled by practitioners without MD's, Engel's intervention had a ready-made constituency in those for whom the radicalism of the antipsychiatrists and the absolutism of the biomedicalists were both unpalatable. Instead of seeking to force psychiatry into the existing medical paradigm, Engel (1992, 320) aimed to use psychiatry's incoherence as a wedge to transform medicine as a whole, by showing that its central commitment to the biomedical model was no more than dogma.

Engel attributed the ideological nature of biomedicine, which he characterized as an allegiance to a reductionist, physicalist treatment of disease states as biological dysfunctions, to broad trends in intellectual history. "With mind-body dualism firmly established under the imprimatur of the Church," he wrote,

> classical science readily fostered the notion of the body as a machine, of disease as the consequence of breakdown of the machine, and of the doctor's task as repair of the machine. Thus, the scientific approach to disease began by focusing in a fractional-analytic way on biological (somatic) processes and ignoring the behavioral and psychosocial. (Engel 1992, 321)

The biopsychosocial model aims to counter this influential philosophical dogma by integrating an understanding of the patient's psychosocial context, including their broader healthcare context. For Engel, this approach was a crucial corrective not just for psychiatry but for medicine as a whole. The exclusion of "mental substance" (or its modern analogs) caused, in his view, a general crisis for not only clinical but also for scientific understanding (Engel 1980, 103). Engel's professional passions, over the course of his career, came to focus on the integration of person-level explanations into our understanding of such quintessentially somatic conditions as heart disease (Ghaemi 2010, 44). As such, his presentation of the biopsychosocial model is primarily addressed to the general physician, and makes the case for treating social factors as relevant to every case of medical decision-making.

Bolton and Gillett agree with Engel's emphasis on the distorting influence of Cartesian dualism, but also agree with critics like Ghaemi who think that his proposed solution—of a general biopsychosocial model—is too vague and unsatisfactory with respect to the scientific details and the philosophical framework (indeed, Ghaemi has argued that the model ultimately has centrifugal, rather than centripetal, effects because of its milquetoast metaphysics, which he believes amounts only to a vapid sort of pluralism). Bolton and Gillett believe, however, that critics err in looking to the model itself to fill in the specifics, which should instead be gathered empirically for each specific stage of each specific health condition. "In this sense", the authors write, "there are multiple specific biopsychosocial models" (Bolton and Gillett 2019, 15); one might think they are too modest here, insofar as their account actually gives rise to *countless* new models! They are quick to correct the idea, however, that they are therefore pushing for (in the language of this paper) centrifugalism, emphasizing that a general model is still needed. Only a unifying framework can provide the "foundational theoretical constructs" that medicine needs. These theoretical constructs, in Bolton and Gillett's view, are "the ontology of the biological, the psychological, and the social—and especially the causal relations within and between these domains" (2019, 19). In other words, they are replacing the "massive historical baggage, carried in the long history of physicalism, dualism and reductionism" with a modern metaphysics that can ground the collected scientific findings of biomedical research. For medical findings, the authors argue, simply *are* biopsychosocial. What will unify medicine, countering the outward push of the vestiges of dualism, is a new theoretical framework recognizing these more inclusive ontological facts, and providing theoretical tools, like a new theory of causation that allows for not only bottom-up but also top-down causation.

Engel was also interested in the role of medical ontology in grounding the biopsychosocial model, and drew on the systems theory in vogue at the time he was writing. For Bolton and Gillett, the new tenets of biopsychosocial causation are to be grounded in modern theories of information-based regulatory control—here they go back to the work of Schrödinger to ground their account of biological systems via an antireductionist biophysics. They also broaden their exploration of top-down causation to include personal agency as a core function of psychology that in turn impacts the biological. The body, therefore, can be "characterised not in mechanical terms, but in terms of functional processes involving information control" (Bolton and Gillett 2019, 79); in the production and management of information, the mental and the physical are "entangled".

I want here to emphasize the close connection in Bolton and Gillett's project between the causes of centrifugalism they attribute to medical theory—physicalism, dualism, and reductionism—and their favored metaphysical counterforce. Like other critics of the biopsychosocial model, the authors emphasize the powerlessness of the model if its content is allowed to be shaped by the weight of a problematic philosophical tradition. When emphasizing that the task of their new model is "defining biopsychosocial ontology and causation," they note

> the special need for this because [of] the deeply entrenched assumptions of physicalism, dualism and reductionism that have been so influential in the development of the life and human sciences. (Bolton and Gillett 2019, 138)

They believe that that tradition has caused medical researchers to neglect the pursuit of certain scientific facts, namely those that require a non-dualistic, non-physicalist, or non-reductive ontology: "With these assumptions, only physical properties and causation appear real, while the mind is a non-causal epiphenomena [sic], and social organization and processes can hardly be comprehended at all" (2019, 138). Accordingly, their project aims to not only provide the missing ontology, but to argue that the biopsychosocial model *must* contain such an undergirding conceptual structure if medical facts are to be legible to scientists.

In the following section I will argue that post-Cartesian philosophy, while a distal cause, is not the most immediate centrifugal pressure on at least one branch of medicine where it is often cited: psychiatry. Engel himself acknowledged the general point, writing, "The power of vested interests, social, political, and economic, are formidable deterrents to any effective assault on biomedical dogmatism" (1992, 328). Bolton and Gillett pay

nuanced and generative attention to the role of autonomy and recognition in the individual's encounter with their social worlds, but they group such factors under the social arm of their unified model, and therefore approach them from a metaphysical perspective. I agree with the authors that the social, political, and economic forces driving biomedicalism are powerful, but I argue in the following section that there is little reason to think they will be attenuated by the introduction of the "right" metaphysics. This is not because most advocates of biomedicine are committed to the view that the mind or social organization and its processes are insubstantial, epiphenomenal, or incomprehensible; it is that they do not find these levels of explanation relevant to medicine's most rewarding projects. After explaining how these non-philosophical forces operate in psychiatry in the following section, I show that while Bolton and Gillett's model can offer a valuable corrective to them, it is ethical counterforces that are more likely to take hold.

3. Centrifugalism in Psychiatry: Other Sources

I have no doubt that philosophical concepts have been crucial to psychiatry's evolving self-image. Alongside the ones that Bolton and Gillett invoke, we can cite the enthusiasm for operationalism in mid-twentieth century philosophy of science that, some believe, entered the psychiatric discourse by way of a talk to the American Psychopathological Association by the logical empiricist Carl Hempel in 1959 (Hempel 1994). This is a case, though, that shows the complexity of establishing philosophical influence; the idea that Hempel caused the APA to immediately pivot to a new approach for the DSM-III has been debunked (Fulford and Sartorius 2009; Schaffner and Tabb 2014; Aragona 2015). Taking this episode as a cautionary tale, Blashfield and Cooper (2018) have argued that philosophers can be lulled into creating origin myths about their own field—philosophy of psychiatry—which in fact exaggerate the influence of philosophy on psychiatry, for the obvious reason that it is validating. At the same time, it is clear that the language of operationalism was taken to be germane both by philosophers and by psychiatrists themselves, such that it was useful as a means of characterizing shifts that were already underway (Tekin 2019). My sense is that something similar has happened with Cartesian dualism, on a grander scale.

In any event, I believe the most significant conceptual vectors of contemporary psychiatry's development to be more recent and more mundane. I will discuss two in this section: intradisciplinary specialization, and market pressures favoring translational research (that is, research that applies basic science findings to medical therapeutics) over clinical

research (that is, original research on human subjects). Each of these vectors has contributed to the contemporary moment, in which the unity of psychiatry's different constituencies—clinicians, researchers, and patients—is at a nadir. The dramatic rift between the National Institute of Mental Health (NIMH) and the American Psychiatric Association (APA) in the early 2000s, brought on by the NIMH's introduction of an alternative to the DSM for researchers, brought these tensions into explicit view. This alternative, the Research Domain Criteria matrix (RDoC), did not aim to replace the DSM in clinical contexts—if it did, it would have been a centripetal force, not a centrifugal one. Rather, the NIMH sought to break what Steven Hyman has called the "epistemic bottleneck" that the clinical conceptual framework imposes on the research setting. Hyman lamented that research questions were neglected when they crosscut the DSM's diagnostic categories, because of the challenge of finding causal mechanisms in heterogenous research samples (Hyman 2010; for discussion see Tabb 2015). When he took over the NIMH's directorship from Hyman in 2002, Thomas Insel (2014) zealously ushered in not only RDoC but also a new vision of psychiatry as "clinical neuroscience".

The introduction of RDoC was significant because it aimed to sever one of the main centripetal forces acting on psychiatry: the hold the DSM had over both clinicians and researchers. The fractious relationship of those working in and around psychiatry to the DSM was already well established. Theorists have noted that clinicians themselves have for decades used the manual less as a scientific guide for understanding psychopathology than as a codebook for managing insurance reimbursements (First and Westen 2007; Whooley 2010). And indeed many clinicians do not *need* a scientific guide; their work is about setting clients up with social services and managing care, including medications which are prescribed on the basis of inductive expertise at best and trial and error at worst. Although it has gotten less attention, it is notable that during the same years RDoC was developed, psychoanalytically-oriented clinicians went so far as to adopt their own manual, the *Psychodynamic Diagnostic Manual,* out of frustration with the DSM. More recently a large coordinated effort to offer a new psychologically-grounded alternative to both the DSM *and* RDoC has taken off, called the Hierarchical Taxonomy of Psychopathology Consortium (HiTOP) (Kotov et al. 2017).

While the NIMH's introduction of RDoC has been taken as a declaration of war against the DSM, this broader context suggests it may go the other way: specialization within the field has made it harder for the DSM, regularly referred to as psychiatry's "bible," to work for everyone amidst the mounting schisms (Lilienfeld 2014). As the complexity of mental illness has emerged with advances not only in the basic sciences but also

in fields like epidemiology, sociology, and human rights, the need for care teams that bring together experts with very little overlap—such as social workers and geneticists—comes ever more into view. The strain put on the DSM to be of use to all these constituencies has been enormous, unparalleled by most other diagnostic instruments (Kutschenko 2011a). Nonetheless, it is hard to imagine something replacing the DSM's crucial role as what Lara Kutschenko Keuck has called an "epistemic hub", facilitating "large-scale interactions without necessarily providing a complete infrastructure". According to Keuck, broadly applied classification systems like the DSM "can be regarded as important nodal points for various actors in biomedical and epidemiological research, clinical practice, and public health" (Kutschenko 2011b, 594). When the hub cracks, the spokes fly loose, and the wheels begin coming off the wagon.

Given all this, the fact that the NIMH decided it advisable, even possible, to do psychiatric research without appeal to the constructs clinicians use to diagnose and treat patients shows how far specialization has come within psychiatry. About the growing gulf between the different constituencies working in and around psychiatry there is much to say, and happily we have historians to say it (see, for example, Halliwell 2013; Menninger and Nemiah 2000; Shorter 1997). From the swelling ranks of case workers to the dwindling ranks of psychoanalysts, the evidence points to these changes being explicable mainly in terms of twentieth-century developments in economics, in labor, and in social policy, rather than as a result of a resurgence of dualist or physicalist commitments. The swing of the pendulum over the course of the twentieth century between the psychoanalytic era's emphasis on early childhood experience, memory, and psychodynamics to the biomedical emphasis on functions, dysfunctions, and physiology does not correspond to any contemporaneous movement in philosophy, whose own "mechanistic revolution" came centuries earlier. Within psychiatry, reductionism—that is, the favoring of explanations that focus on causal relationships between wholes and their constituent parts—was on the rise in psychiatry in the 20th century, but whether it precipitated or resulted from specialization is not obvious. What is clear is that the increasing silos of biomedical research, clinical research, and clinical practice, and the increasing breakdown in interaction between the specialists working in each, has been accompanied by a growing prioritization of basic science and translational research within the field. Biomedicalism is winning.

Members of the American Psychological Association recently sounded the alarm about the NIMH's shift towards "clinical neuroscience" in an open letter to the DSM-5 task force, writing "In light of the growing empirical

evidence that neurobiology does not fully account for the emergence of mental distress, as well as new longitudinal studies revealing long-term hazards of standard neurobiological (psychotropic) treatment, we believe these changes [in favor of biological description] pose substantial risks to patients/clients, practitioners, and the mental health professions in general" (Kamens et al. 2017). The sense among psychologists, social workers, epidemiologists, and other researchers that the NIMH was deprioritizing their research in favor of basic science and translational research has been recently verified empirically (Teachman et al. 2018). These repercussions are rippling far beyond the NIMH itself and other government agencies; for example, Schwartz et al. (2016) note that psychology departments are increasingly changing their names to sound more biological, often by adding the word "neuroscience". Karina Stone and colleagues have demonstrated, using a literature review of articles published in 2008, that about half of all articles in the two major psychiatric journals—*American Journal of Psychiatry* and *The Archives of General Psychiatry*—in that year treat biological themes, as opposed to epidemiological, clinical or review treatment studies (Stone 2012). Strikingly, this percentage was far higher than in leading internal medicine journals, where the number of biologically-oriented papers was only 22%. Psychiatry has become a less hospitable field for those doing clinical, as opposed to biomedical, research.

Bolton and Gillett themselves take an optimistic view of the NIMH's new orientation, suggesting that RDoC could act as a centripetal force insofar as "it could be elaborated in various ways to have broader scope appropriate for the biopsychosocial model" (2019, 128). This sort of elaboration is where their model really shines. By defining the sphere of psychiatry as an entangled systems of regulatory control mechanisms that span a broad scale, from the molecular architecture of organic matter to the individual making choices in response to their environment, Bolton and Gillett show how the limitation of psychiatric inquiry to certain levels of analysis will impoverish the field. Their ambitions for RDoC include the integration of health conditions pertinent to mental functioning, as well as attention to the stages of disease progression and maintenance, and the inclusion of population as opposed to just individual-level information. Their discussion shows how their framework has the potential to guide the expansion of the RDoC matrix beyond its current constructs and domains, which are drawn quite narrowly from cognitive neuroscience. It could give principled grounds for expanding the NIMH's vision of psychiatric research to address the concerns of those researching causal pathways that, while nonbiological, are no less legitimate scientific targets.

Despite the power of Bolton and Gillett's model and the ease with which it could be applied to expand the matrix for future iterations of RDoC, I find it unlikely that the NIMH will be tempted. This is because the NIMH's commitment to reductive explanations does not come from an underexposure to metaphysics, but rather from market pressures that favor certain levels of medical explanation over others. While Bolton and Gillett present RDoC as open to a biopsychosocial approach because of its range of levels of analysis (2019, 126), the highest level of the current matrix is patient self-report—there is no place for social or environmental factors. This is because RDoC was envisioned quite explicitly as psychiatry's debut within the new "precision medicine" paradigm, a hugely influential global push by governments and private research and development institutes to reorient biomedical research towards viable pharmaceutical targets (consider, for example, the title of Insel's 2014 paper in the *American Journal of Psychiatry*, "The NIMH Research Domain Criteria Project: Precision Medicine for Psychiatry"). In line with these broader precision aims, RDoC's architects have stated explicitly, through a series of "postulates," that the matrix is intended to prioritize neurobiological explanations over other levels of analysis:

> First, mental illnesses are presumed to be disorders of brain circuits. Secondly, it is assumed that the tools of clinical neuroscience, including functional neuroimaging, electrophysiology, and new methods for measuring neural connections can be used to identify dysfunction in neural circuits. Third, the RDoC approach presumes that data from genetics research and clinical neuroscience will yield biosignatures that will augment clinical signs and symptoms for the purposes of clinical intervention and management. (Morris and Cuthbert 2012, 33)

Rather than the specter of post-Cartesian thought, I believe that the NIMH's shift towards neuroscience is motivated by the same factors as the shift towards genetic research in precision medicine writ large. The development of psychopharmacology has stalled horribly, and as a result the drug industry has lost interest in researching new treatments for the DSM's diagnoses—they don't pay. The dramatic success of precision medicine drugs in other fields (for example Herceptin, an effective treatment for cancers that are HER2 receptor positive) has revived hope among biomedical researchers that a turn away from signs and symptoms and towards molecular biomarkers will be transformative. About this, too, I am skeptical (see Lemoine and Tabb, forthcoming), but it seems undeniable that the NIMH's attempt to pry biomedical psychiatry free from the conceptual strictures of the clinic follows along from the economic

realities facing its researchers. It seems doubtful a philosophical intervention alone could counter the centrifugal forces of the market, which are pushing clinical research that is deemed profitless to the periphery.

4. The Centripetal Power of Ethical Principles

Building on the previous section, I argue here that if my analysis of psychiatry's current centrifugal pressures is correct, it follows that the best way to address them is not merely through the introduction of a new ontology, but through also making a normative case for the value of such an ontology. I have suggested that the competition for limited resources has driven the split between biomedical psychiatry and clinical psychiatry—the two have been pulled apart not, I have argued, because of entrenched dualism, but because of market forces. There has long been confusion about psychiatry's self-image, with some of its practitioners seeing it as applied neuroscience, some as applied psychology, some as a social welfare project, some as a humanistic quest, etc. But a shortage of resources means that a thousand flowers cannot bloom. While a more inclusive ontology such as that proposed by Bolton and Gillett would refocus psychiatry's scattered attention through its top-down emphasis on the person, its adoption would need to be justified for researchers whose careers have been shaped by centrifugal pressures towards specialization. For many psychiatrists, the disaggregation of biomedical research from clinical practice makes their work possible.

Importantly, such disaggregation is also compatible with a commitment to a fundamentally unified biopsychosocial ontology. Given psychiatry's division of labor, a researcher can recognize the reality of the psychological and social aspects of mental illness but ignore them during a day's work in the lab. In other words, while the biopsychosocial framework seeks to remind biomedicine of its need for psychological and social components on the grounds of ontological entanglement, given the successes of neurobiology in explaining cognition from within a reductionist frame, and the current trends in federal and private funding, this is a hard case to make. Furthermore, while the adoption of a new biopsychosocial ontology would give a rationale for a more evenhanded approach to psychiatric research at the structural level—encouraging funding of both biomedical and psychosocial investigations—handwringing about the exclusion of the psychosocial has not, so far, been effective at countering the powerful centrifugal motion stirred up by increasing investment in the lucrative promise of precision psychiatry.

Using Bolton and Gillett's (2019, 121) language of "modifiable causes", that is, promising targets for intervention, we can say that apologists for the NIMH's neurocentrism are favoring causes operating at the neuroscientific level because they seem the most rewardingly modifiable. Here are Cuthbert and Kozak, for example:

> [I]t is clear that a diagnostic system based upon empirical data from genetics, neurobiology, and behavioral science is desirable to move toward an era of precision medicine where patients are diagnosed and treated according to accurate and appropriately fine-tuned assessments. (Cuthbert and Kozak 2013, 929)

Their emphasis on the applied sciences is pragmatic, not philosophical. It seems the NIMH could very well acknowledge the rich ontology of psychiatry's objects and still insist that some are more worth investigating; the point of RDoC is precisely that biomedical psychiatry does not need clinical psychiatry to point out the appropriate targets for scientific investigation. While Bolton and Gillett are surely right that "it is of fundamental importance in healthcare [that w]e attend to the person, not the body part—and not to psychological signs and symptoms in isolation either" (2019, 116), the fundamental importance of the person to the biomedical researcher is less obvious, given psychiatry's extensive specialization.

As resources shift towards the most powerful interest groups in psychiatry—those with the capital to invest in innovation—and away from those at the less glamorous front lines of mental healthcare (such as social workers, therapists, and general practitioners) there are not only philosophical but practical repercussions. Ethical arguments attending to these repercussions have the potential to bring critical attention from a large range of stakeholders. On ethical grounds one can question whether people's basic rights to healthcare are best served by a psychiatry reconceived as clinical neuroscience (Kirmayer and Crafa, 2014); whether medicine driven by powerful economic interests will align with best bioethical practices (Jeungst et al. 2016); or whether discoveries in neuroscience or genetics, funded by tax-payer dollars, are liable to translate into transformative medical treatments any time soon (Tabb 2020). These questions cannot be brushed aside on the grounds that psychiatric biomedicine is doing just fine without the psychosocial, because they question what "just fine" really amounts to. Questions like these implicate not just to those trying to do good science or provide effective care, but also those who use the mental healthcare system, or even just pay taxes.

Their answers require top-down thinking, not with respect to levels of ontological complexity, but with respect to our higher-order ethical commitments from which decisions about care are deduced. Joseph Margolis has argued that

> medicine is ideology restricted by our sense of the minimal requirements of the functional integrity of the body and mind (health) enabling (prudentially) the characteristic activities and interests of the race to be pursued. (Margolis 1976, 253)

These prudential interests should not, Margolis emphasizes, be confused with the natural functions of the human organism, nor even with the generic values of rational agents. We must attend to the "ulterior goals of given societies" that "reflect the state of the technology, the social expectations, the division of labor, and the environmental condition of those populations" (Margolis 1976, 252). Elsewhere I have argued that while our moral reasoning about such questions relies on empirical facts, it cannot be reduced to them (Tabb 2020). The empirical facts—facts like *how* transformative funds spent on basic research will be to future healthcare advances, or *when* these payoffs will come—rely on our understanding of causes, mechanisms, and systems. But only a broader ethical lens can bring into focus what we should do in response to these facts.

I am not the first to worry that without a unified ethical framework, an expansion of medicine's explanatory projects may only contribute to its dissolution. Moving beyond the case study of psychiatry, in the fractious scholarly debates over the value of precision medicine, critics from a variety of disciplines have expressed worry that the race to disrupt the medical industry with new discoveries can cause resultant healthcare inequities to be obscured. As Ron Bayer and Sandro Galea have written,

> Research undertaken in the name of precision medicine may well open new vistas (…). But the challenge we face to improve population health does not involve the frontiers of science and molecular biology. It entails development of the vision and willingness to address certain persistent social realities, and it requires an unstinting focus on the factors that matter most to the production of population health. (Bayer and Galea 2016)

The payoff for the grinding work of addressing longstanding healthcare inequities and failures in the mental healthcare system is far from

immediate, and therefore the research that would support it is disincentivized within a free market.

Roberto Lewis-Fernández and his coauthors have made similar observations in the context of mental health research, arguing that the shift towards basic and translational research in psychiatry risks neglecting

> thorny details, such as what proportion of the budget should be allocated to what research areas; the near-term public health consequences of particular priorities; and how to leverage inter-agency collaborations to attain a robust and sustainable public health impact. (Lewis-Fernández et al. 2016, 509)

Given that the NIMH is the most significant source of public funds for psychiatric research in the United States, in the American context funding is something of a zero-sum game. In the decade surrounding RDoC's introduction, funding for clinical trials was cut by about a third; the Division of Services and Intervention Research and the Office of Research on Disparities and Global Mental Health was cut by almost 17%; and spending on basic neuroscience went up by 28% (Insel 2015).

What would foundational principles be that could help us navigate these bioethical challenges? They might draw on common understandings of medicine's ultimate aims to give grounds for championing some sorts of medical endeavors over others. A reason to advocate against the centrifugalism of precision medicine, for example, could be that one believes medicine to be more beholden to patients than scientific projects of discovery. Margolis believes medicine to be "primarily an art, and, dependently, a science: it is primarily an *institutionalized service concerned with the care and cure of the ill and the control of disease*" (1976, 242), for which biological understanding is useful but not essential. Under such a view, funding bodies would have an obligation to make sure that any basic science research they fund has clear clinical application. Now of course immediately, longstanding ethical challenges jostle for attention—is it better to deliver imperfect care to patients in need now than to focus on transforming care options for future generations? Does society have an obligation toward the "worried well"—that is, to manage the daily stress of life? Insofar as it can be argued that poverty is a leading cause of mental illness, should the purview of mental health policy extend to questions of social welfare distribution? Etc. Developing worthwhile ethical principles to populate an ethical biopsychosocial model would take the same keen attention to our best bioethics, public policy, and political theory that Bolton and Gillett have paid to our best contemporary theories of causation and ontology.

A more generous metaphysics that includes factors like personal agency is certainly friendlier towards this kind of ethical project than one which dismisses agency as epiphenomenal. But as a unifying framework, the biopsychosocial model has traditionally lacked the specificity to structure these medical-ethical debates. In other words, it has failed to provide an account of Margolis' prudential functions, those capacities that we prioritize not because they are natural to us but because they allow us to live in the ways we deem right. Whether to prioritize resolving Lewis-Fernandez et al.'s "thorny details" or instead to attend to the fascinating puzzles of basic neuroscience or behavioral genetics cannot be answered on the basis of a pluralist ontology alone. Insofar as the whole person—from genes to environmental interactions—is implicated in these questions, the biopsychosocial model offers no grounds for resolution. However, Bolton and Gillett argue explicitly that their model also holds a place for ethics within its ontology, in so far as it follows from agency being "thoroughly biological" (138) that it "becomes involved with morality," due to the entanglement of the biological, psychological, and social (88). Before closing I want to consider whether the theoretical ethical principles I am looking for "fall out" of their model in some way that would render the addendum I am proposing unnecessary.

5. The Normativity of the Biopsychosocial Model

Seeing RDoC as a wedge to move the basic and translational sciences towards the core of the discipline can explain why its advocates have ignored another repercussion of their attempted coup against the DSM: the loss of a bellwether for distinguishing the normal from the pathological. The architects of RDoC have shown little interest in taking up the mantel, emphasizing that they are merely interested in the elucidation of mechanisms, not in the demarcation of disease categories. But which mechanisms count as psychiatric? This is not just about semantics; the NIMH's mission is to fund research into mental health, not physiology, and RDoC is to a large degree about shaping what research counts as what (Tabb 2020). Without some grounds for ruling on what counts as psychiatric and what doesn't, the NIMH can increasingly fund basic research in, e.g., neuroscience or genetics, moving the institute ever further away from its traditional focus on mental illness as a societal problem (Bloom 2002, 165).

Insel, writing with Bruce Cuthbert, has suggested that maybe mental disorders can be defined as extremes of functional variation, writing,

> The idea [of RDoC] is to start by specifying basic dimensions of functioning, and their implementing brain circuits, that have been identified by the last several decades of research in brain and behavior. Then, in this light, mental disorders are considered as extremes at one or both tails of those normal distributions. (Cuthbert and Insel, 2010, 312)

This approach to delineating diseases—as tails on a normal distribution—is profoundly unsatisfactory, as philosophers of medicine have long pointed out (Boorse 2011, 21). Which tail (one or both)? Where is the cut-off (and who decides)? Jerome Wakefield has described RDoC's naïve approach to the demarcation problem as a failure of conceptual validity. "Whatever its errors," Wakefield writes, the DSM

> remains an attempt to delineate the domain of psychological conditions that fall under the concept of disorder. RDoC offers nothing to replace the [DSM's] efforts to delineate the domain of disorders and provide a target at which construct validation can aim. (Wakefield 2014, 38)

The results are "so weak that it is difficult to envision success" (ibid.).

Broadly speaking, attempts by philosophers and psychiatrists to provide an analysis of mental disorder that could help demarcate psychiatry's objects have been copious, heated, and ultimately inconclusive (for recent moves in this debate see Faucher and Forest 2021; for a critical analysis of it see Lemoine 2013). Bolton and Gillett themselves offer a hybrid view, combining naturalist and normativist elements, in which they argue that normativity "is fundamental to biological regulatory control mechanisms" (2019, 68) and that therefore disease can be understood, generally, in terms of failures of function produced by these feedback mechanisms. They suggest that the levels of dysfunction where mental pathology manifests in practice—the psychological and the social—are emergent manifestations of these biological dysfunctions (2019, 72). However, on the grounds of their comfort with top-down causation, they also suggest that dysfunction can be located in any part of a system that is both modifiable and the cause of error: "From this point of view, dysfunction attribution is in part—and somewhat paradoxically—shorthand for belief about promising possibilities for change" (2019, 121). The *need* to change, they suggest elsewhere, comes with patients' self-report of "distress: with worry and fear about their safety and their future and their dependents" (135).

Demarcations between the normal and the pathological that rely even in part on naturalist theories of dysfunction have, to my mind, been

convincingly problematized by philosophers like Ron Amundson, who have argued that the ontological makeup of the individual organism can shed light on *mode* of function, but not establish *level* of function. Writing in the context of disability, Amundson argues that what matters for defining disability is an individual's capacities within a given environment; their functional makeup is irrelevant to determinations of health. "If we thought merely about *level* of functional performance, rather than mode, fashion, or style of function," Amundson writes, "the disadvantages of disability would not seem so natural and inevitable" (2000, 48). Amundson's case for rejecting biological theories of dysfunction is also an ethical one—to focus on mode is to facilitate the continuation of historic abuses against those who function differently.

Bolton and Gillett recognize disabilities as "a special case" due to the lack of modifiable causes within organism's system, allowing that here errors "can be legitimately attributed to (…) external factors" (114). While they insist that "disability related concepts and practices involve a complex range of and interaction between biological, psychological, social, moral and policy factors" and therefore "cannot be so much as articulated without a full biopsychosocial framework," it is unclear on what grounds their new ontology—reliant as it is on locating dysfunction *within* the system—could offer robust support to a social model of disability like Amundson's, which takes the black-boxing of function, and a turn to the disabling features of the environment, to be an ethical imperative. At one point in their book, Bolton and Gillett seem to accept that while generally they are committed to locating "the problem—the dysfunction—in the person", they must make an exception for conditions that are lifelong and/or not amenable to change (2019, 120).

The fact that the new biopsychosocial framework has little to offer on these conditions should give us pause, given the percentage of mental disorders that display them. Furthermore, those diagnosed with psychiatric disorders are increasingly conceptualizing their conditions in terms of difference rather than dysfunction, in alignment with the social model of disability. While there has always been robust activism in response to the perceived overreach of biomedical psychiatry, contemporary activists have introduced a new conceptual framework for thinking about this resistance. Instead of denying that purported mental illnesses have any clinical relevance, like the radical antipsychiatrists of 1960s and 70s, some contemporary critics argue for destigmatization alongside new demands for healthcare justice. To advocate for neurodiversity is to believe that healthcare, social services, and culture broadly construed must change to offer a broader range of supports, allowing not only the neurotypical but the neurodiverse to flourish. To be neurotypical, in other words, is just to

have the sort of psychological profile that is already served (more or less) well by one's environment, and there is no reason to see such a profile as innately healthier, rather than just more convenient under the current circumstances. Given the neurodiversity movement's suspicion of essentializing ontologies, its reliance on social constructionist narratives of illness, and its impatience with biomedical levels of description, its best ontological allies may be quietist, not pluralist. What would really help is a psychiatric ethics capable of justifying their claim to healthcare as a human right, even in the absence of dysfunction.

6. Final thoughts

I have argued that the centrifugal forces causing rends in psychiatry's conceptual fabric are due to a confluence of political, economic, and cultural factors. The displacement of the DSM as the field's arbiter of the normal and the pathological was both a result and a driver of increased specialization within the field, which led to new antagonisms and struggles. The economic promise of the precision medicine model, which matches patients with novel therapies on the basis of biomarker testing, has caused an influx of financial support for biomedical approaches to psychopathology. Advocates of precision psychiatry need not deny that there are other levels on which psychopathological phenomena can be found, and intervened upon—such as the psychological or the social. But they may doubt that there are modifiable causes to be found at these levels, or that these causes are as *rewardingly* modifiable as those found at the level of the neural circuit. While Engel wished psychiatrists to be "concerned primarily with the study of man and the human condition" (1992, 327), this hardly seems realistic for the twenty-first century biomedical researcher, whose lab work in psychiatric genetics or in neuroscience may never require meeting a patient.

The result of this recent enthusiasm for precision psychiatry is that the field is increasingly pulled in different directions. Its practitioners rely on traditional disease categories as well as their own expert knowledge of psychopathology to do their work, while its researchers borrow the concepts and methods of the basic sciences for theirs. Similar changes are underway in other fields where the precision paradigm has taken hold. To counter this centrifugal motion, I have suggested, a new ontology is not enough, because the motivations for the split do not result from monist or reductionistic ontological commitments as much as they do from economic and political factors. These systemic pressures on the profession force different sorts of practitioners farther apart, and reward psychiatric research that diffuses its center of gravity away from immediate mental

health crises. Accordingly, to convince the diverse stakeholders in psychiatry that it is important to all work toward the same thing, ethical arguments hold greater promise. They can exert pressure on the powers making decisions about what kind of psychiatric research is worth funding, and what kind of mental healthcare is worth expanding. A new ontology that takes seriously the complex feedback loops between the biological, the psychological, and the social has the potential to encourage a revaluing of neglected populations. But the need to adopt such an ontology may only become clear when it is shown how the exclusion of psychosocial dimensions causes us to fail in our ethical obligations.

It is worth noting that the biopsychosocial model itself might be conceived of in purely prudential terms, instead of in metaphysical terms. Such a theory would offer a model of psychiatry as unified by the biological, psychological, and social aspects of people's mental health, not because these are aspects of a unified ontology, but because they form a unified set of obligations. In her "Neurodiversity at Work: A Biopsychosocial Model and the Impact on Working Adults," Nancy Doyle notes that the biopsychosocial model can be maintained even amidst "ontological controversy" over the nature of mental illnesses like autism. She glosses its biological component as "therapeutic intervention" rather than as referring to any (dys)function within the individual, and the model as a whole is taken as a pragmatic one, with the explicit aim of realizing the best outcomes for neurodiverse people in the workplace (119). By dismissing concerns about the place of the pathological, however, this account is ultimately centrifugal, disaggregating the question of how neurodiverse people should be treated in the workplace from larger ones concerning psychiatry's biomedical projects.

In contrast, Bolton and Gillett's new biopsychosocial model is exciting for its stout centripedalism, which could ground an ethical framework for *all* of medicine. Yet as it currently stands, the model does not contain foundational principles capable of negotiating, on ethical grounds, between those advocating for biological, psychological, or social approaches to disease. It is this nonpartisan tendency of the biopsychosocial model that has, I think, frustrated critics. This reflects a broader suspicion about pluralism: that one can end up with a conglomerate of models that, taken together, are like the map in Borges' story "On Rigor in Science". Cartographers render this map so exact that it papers over the whole land, rendering itself useless. One feels for Bolton and Gillett when their amendments of the RDoC matrix cause it to grow rather threateningly, in their words, into "a multidimensional monster grid" (130). The authors encourage us to see this complexity and uncertainty as a result of the science itself, rather than the model—"no point in blaming the messenger"

(132). But the moral of Borges's story is that the main responsibility of the modeler lies precisely in picking the right scale for the job. A scientific theory is, in this analogy, not the messenger but the message itself, which aims to render legible the complexity of the modeled system. Which "aetiology of small effect" (132) we take as definitional of health conditions must be made not only by "doing science" but also by making choices between modifiable causes. As Bolton and Gillett note, medicine is an "applied science, seeking to change things, for the better" (2019, 121). If so, the explanatory choices that result from a model should be normative. Determining what differences in function are appropriate targets for medical intervention and which are better left for scientific or societal interventions cannot be read off the individual's own state of functioning or agential status. It relies on broader societal norms concerning well-being, and the ethical commitments of medicine itself.

I believe that general ethical principles could be addended to the new biopsychosocial model without requiring it to give up its neutrality with respect to the relative value of the biological, the psychological, and the social. Instead, the framework could host a normative pluralism analogous to the ontological pluralism undergirding the "multiple specific biopsychosocial models" that Bolton and Gillett allow for, in which the relevance of each aspect will change depending on prudential functions relevant to the case at hand. At the same time, the model could seek to supply the abstract theoretical constructs necessary for a powerful new medical ethics. Being integrated into the new biopsychosocial framework would assure that these theoretical constructs would guide all research and practice falling under the broad reach of the model.

REFERENCES

Aragona, Massimiliano. 2013. 'Neopositivism and the DSM Psychiatric Classification: An Epistemological History'. Part 1: Theoretical Comparison'. *History of Psychiatry*, 24: 166-179.

Bloom, Samuel William. 2002. *Word as Scalpel*. Oxford: Oxford University Press.

Bolton, Derek, and Grant Gillett. 2020. *The Biopsychosocial Model of Health and Disease: New Philosophical and Scientific Developments*. Cham, Switzerland: Palgrave Pivot.

Cooper, Rachel, and Roger Blashfield. 2018. 'The Myth of Hempel and the DSM-III'. *Studies in the History and Philosophy of Biology: Biomedical Sciences.* 70: 10-19.

Cuthbert, Bruce, and Thomas Insel. 2010. 'The Data of Diagnosis: New Approaches To Psychiatric Classification.' *Psychiatry*, 73 (4): 311-314.

Cuthbert, Bruce, and Michael J. Kozak. 2013. 'Constructing Constructs for Psychopathology: The NIMH Research Domain Criteria'. *Journal of Abnormal Psychology* 122 (3): 928-937.

Engel, George. 1992 [1977]. 'The Need for a New Medical Model: A Challenge for Biomedicine'. Reprinted in *Family Systems Medicine*, 10 (3): 317-331.

First, Michael B, and Drew I Westen. 2007. 'Classification for Clinical Practice: How To Make ICD And DSM Better Able To Serve Clinicians'. *International Review of Psychiatry* 19: 473–481.

Fulford, K. W. M., and Norman Sartorius. 2009. 'The Secret History of ICD and the Hidden Future Of DSM'. In *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives,* edited by Matthew Broome and Lisa Bortolotti, 151-158. Oxford: Oxford University Press.

Galea, Sandro and Ronald Bayer. 2016. 'The Precision Medicine Chimera'. *Project Syndicate*. January 14.

Ghaemi, S. Nassir. 2010. *The Rise and Fall of the Biopsychosocial Model: Reconciling Art & Science in Psychiatry*. Baltimore: Johns Hopkins University Press.

Halliwell, Martin. 2013. *Therapeutic Revolutions: Medicine, Psychiatry, and American Culture, 1945-1970*. Rutgers University Press.

Hempel, Carl. 1994. Fundamentals of Taxonomy. In *Philosophical Perspectives on Psychiatric Diagnosis Classification*, edited by John Z. Sadler, Osborne P. Wiggins, & Michael A. Schwartz, 315–332. Baltimore: Johns Hopkins University Press.

Hyman, Steven E. 2010. 'The Diagnosis of Mental Disorders: The Problem of Reification'. *Annual Review of Clinical Psychology* 6: 155-179.

Insel, Thomas R. 2015. "Anatomy of NIMH Funding," available at http://www.nimh.nih.gov/funding/fundingstrategy-for-research-grants/the-anatomy-of-nimh-funding.shtml.

Insel, Thomas R. 2014. 'The NIMH Research Domain Criteria (RDoC) Project: Precision Medicine for Psychiatry'. *American Journal of Psychiatry* 171 (4): 395-397.

Insel, Thomas R., and Remi Quirion. 2005. 'Psychiatry as a Clinical Neuroscience Discipline'. *Journal of the American Medical Association* 294 (17): 2221–24.

Juengst, Eric, Michelle L McGowan, Jennifer R. Fishman, and Richard A. Settersten Jr. 2016. From "Personalized" to "Precision" Medicine: The Ethical and Social Implications of Rhetorical

Reform in Genomic Medicine. *Hastings Center Report* 46: 21-33.

Kamens, Sarah R., David N. Elkins, and Brent Dean Robbins. 2017. 'Open Letter to the DSM-5', available online at: https://www.ipetitions.com/petition/dsm5/.

Kirmayer, Laurence J., and Daina Crafa. 2014. 'What Kind of Science for Psychiatry?'. *Frontiers in Human Neuroscience* 12: 1-12.

Kotov, Roman, Robert F. Krueger, David Watson, et al. 2017. The Hierarchical Taxonomy of Psychopathology (Hitop): A Dimensional Alternative to Traditional Nosologies. *Journal of Abnormal Psychology*, 126 (4): 454-477.

Kutschenko, Lara K. 2011a. 'In Quest of 'Good' Medical Classification Systems'. *Medicine Studies* 3: 53-70.

Kutschenko, Lara K. 2011b. 'How to Make Sense of Broadly Applied Medical Classification Systems: Introducing Epistemic Hubs'. *History and Philosophy of the Life Sciences Part C,* 33: 583-602.

Lilienfeld, Scott. 2014 'DSM-5: Centripetal Scientific and Centrifugal Antiscientific Forces'. *Clinical Psychology: Science and Practice* 21 (3): 269-279.

Margolis, Joseph. 1976. 'The Concept of Disease'. *Journal of Medicine and Philosophy* 1 (3): 238–255.

Morris, Sarah E, and Bruce N. Cuthbert. 2012. 'Research Domain Criteria: Cognitive Systems, Neural Circuits, and Dimensions of Behavior.' *Dialogues in Clinical Neuroscience* 14 (1).

Schaffner, Kenneth F. 1994. 'Psychiatry and Molecular Biology: Reductionistic Approaches to Schizophrenia'. In *Philosophical Perspectives on Psychiatric Diagnostic Classification,* edited by John Z. Sadler, Osborne P. Wiggins, & Michael A. Schwartz. Baltimore: Johns Hopkins University Press. 279–294.

Schaffner Kenneth F., and Kathryn Tabb. 2014. 'Response to Josef Parnas'. In *Perspectives in Philosophy and Psychiatry III: The Nature and Sources of Historical Change,* edited by Kenneth S. Kendler and Josef Parnas. Oxford: Oxford University Press, 213-220.

Schwartz, Seth J., Lilienfeld, Scott O., Meca, Alan, and Katheryn C. Sauvigné. 2016. 'The Role of Neuroscience Within Psychology: A Call for Inclusiveness Over Exclusiveness'. The American Psychologist 71, 52–70.

Shorter, Edward. 1997. *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac.* John Wiley & Sons.

Stone, Karina, Elizabeth A. Whitham, and S. Nassir Ghaemi. 2012. 'A Comparison of Psychiatry and Internal Medicine: A Bibliometric Study'. *Academic Psychiatry* 36 (2): 129–32.

Tabb, Kathryn. 2015. 'Psychiatric Progress and the Assumption of Diagnostic Discrimination'. *Philosophy of Science* 82 (5): 1047-1058.

Tabb, Kathryn. 2020. 'Should Psychiatry Be Precise? Reduction, Big Data, and Nosological Revision in Mental Health Research'. In *Levels of Analysis in Psychopathology* edited by Kenneth S. Kendler, Josef Parnas, and Peter Zachar. Cambridge: Cambridge University Press, 308-334.

Teachman Bethany A., Dean McKay, Deanna M. Barch, Mitchell J. Prinstein, Steven D. Hollon, and Dianne L. Chambless. 2019. 'How Psychosocial Research Can Help the National Institute of Mental Health Achieve Its Grand Challenge to Reduce the Burden of Mental Illnesses and Psychological Disorders'. *The American Psychologist* 74 (4): 415-431.

Tekin, Şerife. 2019. 'The Missing Self in Scientific Psychiatry'. *Synthese* 196 (6): 2197–2215.

Whooley, Owen. 2010. 'Diagnostic Ambivalence: Psychiatric Workarounds and the Diagnostic and Statistical Manual of Mental Disorders'. *Sociology of Health & Illness* 32 (3): 452-69.

EuJAP | Vol. 17 | No. 2 | 2021
UDC: 179:61
https://doi.org/10.31820/ejap.17.2.5

HOW TO BE A HOLIST WHO REJECTS THE BIOPSYCHOSOCIAL MODEL

Diane O'Leary[1]

[1]Center for Philosophy of Science, University of Pittsburgh

Original scientific article – Received: 27/02/2021 Accepted: 27/05/2021

ABSTRACT

After nearly fifty years of mea culpas and explanatory additions, the biopsychosocial model is no closer to a life of its own. Bolton and Gillett give it a strong philosophical boost in The Biopsychosocial Model of Health and Disease, but they overlook the model's deeply inconsistent position on dualism. Moreover, because metaphysical confusion has clinical ramifications in medicine, their solution sidesteps the model's most pressing clinical faults. But the news is not all bad. We can maintain the merits of holism as we let go of the inchoate bag of platitudes that is the biopsychosocial model. We can accept holism as the metaphysical open door that it is, just a willingness to recognize the reality of human experience, and the sense in which that reality forces medicine to address biological, psychological, and social aspects of health. This allows us to finally characterize Engel's driving idea in accurate philosophical terms, as acceptance of (phenomenal) consciousness in the context of medical science. This will not entirely pin down medicine's stance on dualism, but it will position it clearly enough to readily improve patient care.

Keywords: *Biopsychosocial model; holism; dualism; philosophy of medicine; psychosomatic medicine*

Correspondence: drdianeoleary@gmail.com

1. Introduction

The biopsychosocial model (BPSM) has two central problems: one philosophical and one clinical. First, while the model turns away from reductive physicalism, proposing an alternative that brings subjective experience into the scope of medical science, its ontological position is, at best, unclear and, at worst, incoherent. Second, while the model demands a radical change in everyday practice—again, a broadening that will range over not only biological, but also psychological and social considerations––it fails to provide guidance as to what, exactly, a clinician should do to practice in a biopsychosocial way.

In *The Biopsychosocial Model of Health and Disease*, Bolton and Gillett offer a convincing presentation of the BPSM, highlighting these fundamental problems in their own terms, then they set out to resolve them. The result, they suggest, is a BPSM rethought and reinvigorated, one with far more substantial ties to philosophy. The need for this kind of rethinking is very real, as the BPSM has become a kind of dogma for medicine, even if only in marketing, while its shortcomings remain severe. As Bolton and Gillett aptly put it, the result is a crisis for medicine's foundations, one long in the making.

Engel could not have hoped for a more enthusiastic effort at redemption, nearly fifty years into medicine's biopsychosocial journey, and in many ways the effort is invaluable, even ingenious. Where Engel was vague (to put it kindly) about causal connections, Bolton and Gillett fill in the gaps, and in a way that brings the BPSM into current philosophical focus. Most valuable, I think, is their discussion of embodied cognition as a tool for fleshing out the scientific meaning of slogans like "mind-body integration". More than that, authors provide a detailed and wide-ranging account of the kind of complex causal interdependence that can make the BPSM work as a matter of science. Even if we find fault with their account and its idiosyncrasies, its value will remain. The BPSM is so often framed as medicine's softer side, while the evidence-based model fills the slot for hard science. That understanding is a mistake, and Bolton and Gillett will have made that clear even if their particular account of the science can be challenged.

The BPSM, however, is not redeemed by this ingenuity. Philosophically speaking, while Bolton and Gillett devote most of the book to the intricacies of their causal picture across the biopsychosocial spectrum, the model's most glaring, and most pressing, ontological failures are not recognized. Moreover, because medicine's metaphysical confusions have powerful clinical ramifications, Bolton and Gillett's solution to the clinical

problem also sidesteps the BPSM's most pressing faults. I will address each of these issues in turn.

In the end of the day, I will not suggest that Bolton and Gillett's efforts have been wasted. I will suggest that they've been wasted on the BPSM. Nassir Ghaemi (2010, 213) is right, I think, that the BPSM is more a slogan than a model, and we've spent almost fifty years tacking on *mea culpas* and explanatory additions. None of these has begun to give the thing life as a model, because none have addressed, or could address, the radical inconsistencies that have grown out of Engel's original philosophical confusions. But the news is not all bad. There is no reason why we cannot begin anew with a form of holism that takes what works from Engel and lets go of what fails. There is no reason why we cannot, from a clean slate, build a new model for holism that is philosophically sound, scientifically substantial and, above all, optimal for patient care.

2. The Philosophical Problem

Philosophically speaking, the simplest and most salient feature of the BPSM is an ontological expansion of medicine's conceptual foundations. Whatever else we might say about the model as Engel presented it, it is clear that, according to the BPSM, traditional medicine's exclusive focus on the physical body is misguided. To improve things, medicine must expand to recognize the inextricable place for mind, for experience, in the health of the whole person.

From the perspective of current philosophy of mind, this idea is uncomplicated. It is a rejection of reductive physicalism in favor of some form of property dualism or nonreductive physicalism. Practically speaking, however—and in spite abundant research in philosophy since Engel's time on alternatives to reductive physicalism—medicine's conceptual foundations were not clarified by the BPSM. They were confused to an extent that the model itself cannot remedy.

First, there is deep, pervasive inconsistency about the BPSM's most basic ontological position—that is, its position on dualism (O'Leary 2020). On one hand, in the simplest and most obvious terms, many in philosophy of medicine understand the model to be dualistic. For example, Marcum suggests, citing Foss (2002), that "biomedicine is composed of a metaphysical position best defined as mechanistic monism", while "the biomedical worldview is modified in humane medicine with a metaphysical position that is generally dualistic" (Marcum 2008, 394-95). Borrell-Carrio and colleagues see a similar picture in their twenty-five-

year retrospective on the BPSM, concluding that "George Engel formulated the biopsychosocial model as a dynamic, interactional, but dualistic view of human experience" (Borrell-Carrio 2004, 581).

On the other hand, in the borderlands between medicine and psychiatry, the BPSM is generally assumed to be defined by rejection of dualism. In "The persistence of mind-brain dualism in psychiatric reasoning about clinical scenarios", for example, Miresco and Kirmayer explain that "Despite attempts in psychiatry to adopt an integrative biopsychosocial model (…) psychiatrists continue to operate according to a mind-brain dichotomy" (Miresco and Kirmayer 2006, 913). More than that, they define dualism as "the idea that the mind is somehow distinct from the brain and that its essence cannot be reduced to purely material and deterministic neurological mechanisms" (Mireseco and Kirmayer 2006, 913). For those who see the model from this perspective, BPS ontology is characterized by opposition to dualism, by the idea that mind can "be reduced to purely material and deterministic neurological mechanisms".

Though Bolton and Gillett very clearly understand dualism as a problem to be overcome, and a problem that they do overcome with a "new post-dualist framework", the book provides no definition of dualism, no acknowledgement of the common perception that the BPSM is dualistic, and no effort to explain why that perception might be mistaken.

Second, because inconsistency about dualism poses such a decisive threat to the coherence of the BPSM, we must investigate whether it can be understood in a way that accommodates both perspectives. Is it possible for one medical model to both accept and reject dualism? Perhaps, if it accepts one form of dualism while it rejects another, but a picture of that kind would require a clear and well-defined account of its position. Do we find such an account in Engel? Definitely not. In fact, when we take a closer look at Engel's original characterization of the biomedical model, we can actually see how we've ended up with such deep ontological confusion. Engel straightforwardly insisted—not once, but consistently in all of his writings—that

> the biomedical model embraces both reductionism, the philosophic view that complex phenomena are ultimately derived from a single primary principle, and mind-body dualism, the doctrine that separates the mental from the somatic. (Engel 1977, 130)

This, unequivocally, is the malady that Engel sets out to remedy with the BPSM: not reductionism on its own, but reductionism in combination with dualism.

Broadly speaking, these are diametrically opposed views. In the broadest, most unrefined sense, reductive physicalism and Cartesian dualism are mutually exclusive, so it's not possible for Engel to be correct in framing the BMM as reductive dualism, or dualistic reductionism. In the broadest sense, then, the BPSM is aiming for an incoherent goal, setting out to reverse a position that was impossible in the first place.[1]

Of course as proponents of the BPSM, we could take a more refined view of our ontological options. We could position ourselves between the poles of reductionism and Cartesian dualism with some form of property dualism, for example. Such a position would be a fine antidote to both of those polarities—but again, this would require quite a lot of philosophical refinement. We'd need to clarify, as Susan Schneider does, that while

> contemporary philosophy of mind sees the question of the nature of substance as being settled in favor of the physicalist (…) dualism about properties, by contrast, is regarded as being a live option. (Schneider 2012, 51)

We'd need an explanation of the difference between Cartesian realism about minds and current realism about mental properties. Then we'd need a discussion of the difference between nonreductive physicalism (where we accept that mental properties are distinct from physical properties, but reject dualism), and naturalistic dualism (where we accept that mental properties are distinct from physical properties and accept dualism).

Does Engel provide an account of this kind, where we can make sense of the model's contradictory views on dualism through a more contemporary, more refined account of nonreductive alternatives? No, though these options really had not been laid out in clear terms when Engel was formulating the BPSM. Do we get an account of this kind in the "biopsychosocial ontology" that Bolton and Gillett promise to provide? Still, no. In fact, Bolton and Gillett fail to mention property dualism even once. In the brief passage that mentions nonreductive physicalism, they

[1] Bolton and Gillett eloquently explain that "physicalism and dualism are twins, one born straight after the other, combative from the start, each refuting the other, the one supported by the great edifice of modern mechanics, the other known immediately by experience, battling ever since" (Bolton and Gillett 2019, 27). Unfortunately, while they often describe the pairing in the BMM as "physicalist reductionism aided by dualism", they do not explain how it might be possible to hold both positions simultaneously.

dismiss the view, inexplicably, as a "purely 'metaphysical' doctrine", one that "probably has given up on being much or anything to do with the sciences" (Bolton and Gillett 2019, 161).

Third, we have been unable to resolve the BPSM's ontological inconsistency because the term 'dualism' has been defined in a way that makes philosophical clarification impossible. This problem can be traced directly from Engel to Bolton and Gillett.

The only way to make sense of the idea that reductionism and Cartesian dualism go hand and hand is to fudge the definition of dualism a bit. For Engel, as for his colleagues, as for most of those who've worked with the BPSM for the last forty years, dualism is not an ontological position, not a view on how many kinds of substances or properties exist. Engel's brand of dualism is an epistemological position, a choice each of us can make in our thinking. When we separate mind and body in our thinking, we are dualists, and when we integrate them, we defeat dualism. Unfortunately, dualism is actually not an epistemological position. Dualism does not come and go depending on the ideas we prefer or the words we choose. If the world is dualistic, then two kinds of things exist in the world, no matter what we say or think or do in medical practice.

Bolton and Gillett's book is a productive example of this confusion and its catastrophic impact on medicine's foundational clarity. Though authors promise at the start to provide a new ontology for the BPSM, and later they take themselves to have made good on that promise, like Engel, they pair dualism with reductionism, almost as a habit. Like Engel, they feel sure they've conquered dualism "when physical and mental health conditions are brought together (…) rather than being axiomatically separate" (Bolton and Gillett 2019, 109). Moreover, because, like Engel, they believe we settle the question of dualism when we choose not to separate mind and body in our language or practice, they entirely overlook the actual question of dualism, that is, the question of whether minds, or mental properties, exist.

It's important to be clear about why it's philosophically problematic to define dualism as separation of mind and body in our thinking rather than as the existence of minds or mental properties. After all, dualists always do separate mind and body, so it will work out just fine to define it that way as long as we're affirming dualism. The trouble arises when we reject dualism—because we can choose to reject separation of mind and body in our thinking as dualists, or as monists. Marcum (2008) and Borrell-Carrio et al. (2004), for example, both insist that while the BPSM is a dualistic model, one that recognizes both mind and body, it also demands that we

recognize them as unified, rather than separated, in the whole person. Miresco and Kirmayer (2006), on the other hand, insist that the BPSM is a monistic model. From their perspective, it's a mistake to separate mind and body because all the world is physical.

This is the source of the BPSM's philosophical incoherence. We cannot begin to determine whether medicine is or is not dualistic unless we're clear what that question means: does medicine's understanding of health and healthcare require the existence of minds or, alternatively, mental properties? Once we're clear about that, nonreductive physicalism and naturalistic dualism become instant candidates for holism's ontological foundation. While it's certainly possible to argue that both fail to make sense of the whole person in the way that Engel intended, or the way that medicine actually requires, these are the most widely accepted ways to make sense of a holistic vision in contemporary philosophy of mind. We cannot sort out medicine's ontological foundations without considering them.

Admirable as Bolton and Gillett's picture of BPS causes may be, it will not stand as an account of BPS ontology until authors make direct use of it to resolve the BPSM's pervasive inconsistency about dualism. To do so they'd need to recognize that, in the twenty-first century, the question of dualism is serious and meaningful, especially for medicine. It is the hard problem of accounting for the reality of experience in the context of science (Chalmers 1995). More than that, they'd need to acknowledge that, like Engel, they do help themselves to the reality of experience as central to a sound understanding of health and healthcare.

Fourth and finally, any effort to provide a workable ontology for the BPSM must address incoherence in its central claims about mind and body.

> (a) The first step and most important step toward an ontologically coherent picture of the BPSM is to clarify a consistent definition of dualism within the terrain that characterizes contemporary philosophy of mind. That, on its own, would be a monumental accomplishment for philosophy of medicine, one that would reverberate productively through all the medical professions.
>
> (b) Second, we need an explanation of why medicine should reject dualism, if, in fact, it should—because rejection of dualism does not go without saying in philosophy of mind, surprising as that may be to many in the medical professions. Because the question on the table in philosophy is about

> property dualism rather than substance dualism (generally speaking), and, generally speaking, philosophy of mind has accepted the reality of mental properties, rejection of dualism does now require clarification and support. In any area of discourse that depends on recognition of experience *qua* experience, as the BPSM certainly does, it is absurd to proceed as if rejection of dualism goes without saying.
>
> (c) Third, because separating mind and body certainly does not make us dualists, not in philosophy of mind, we need a discussion of the merits and drawbacks of separating them in medicine. The fact is that, by and large, philosophers of mind are comfortable distinguishing mental properties from physical properties. To put that a different way, by and large, philosophy of mind has accepted a real distinction between experiences and the brain states with which they're correlated. "Separation of mind and body", is not a problem in philosophy, at least not *prima facie*. If we want to propose that it's a problem for medicine, either metaphysically or clinically, that idea that will require clarification and support.

While it is certainly possible to address these three issues, it is hard to imagine any way that we might institute revisions on these points in everyday thinking about the BPSM in medicine, psychiatry or bioethics. After fifty years of incoherent wrangling about mind and body, that is to say, the BPSM has come to be defined by its entrenched philosophical inconsistency. Though we surely can repair medicine's conceptual foundations, we will need to see the result as an alternative form of holism, a better form of holism than what we get with the BPSM. I will make some broad points about that project in Part 4, but first it's important to track the BPSM's ontological confusion as it actually plays out at the level of clinical practice.

3. The Clinical Problem

In addition to the formidable challenge of ontological incoherence, the BPSM also faces a practical challenge, that it "lacks specific content, is too general and vague" (Bolton and Gillett 2019, 29) at the level of clinical application. Ghaemi suggests that while the addition of psychological and social considerations do provide greater freedom and complexity in diagnosis and treatment

> [t]his eclectic freedom borders on anarchy: one can emphasise the 'bio' if one wishes, or the 'psycho' (…) or the 'social'. But there is no rationale why one heads in one direction or the other: by going to a restaurant and getting a list of ingredients, rather than a recipe, one can put it all together however one likes. (Ghaemi 2009, 3)

The new options are certainly reasonable (maybe reasonable enough to be obvious for psychiatry), but they're not useful without general guidance as to how they should be used.

Bolton and Gillett propose that this problem can be resolved at the level of research, where new evidence for the relevance of psychosocial factors in specific conditions has now been developed. Clinicians can do without general principles for choosing between bio, psycho, and social options, they suggest. BPS practice can be accomplished purely by applying information from research about specific psychosocial factors for specific conditions. This approach goes a long way toward aligning the BPSM with evidence-based medicine, and I am very much in favor of that kind of effort. In the process, however, it overlooks Engel's vision for BPS practice, the risk it creates in providing diagnostic options without diagnostic guidance, and the sense in which that gap has been filled by ontological confusion.

First, discourse about the BPSM, including Bolton and Gillette's, often fails to appreciate Engel's rich picture of the clinical interview. In "How much longer must medicine's science be bound by a seventeenth century world view?" Engel directly opposes the idea that the clinical relevance of the BPSM could play out purely through the application of research, and his arguments on this point may be the most convincing we find in his work. He explains in detail exactly how the clinical interview is a "means of data collection and processing" (Engel 1992, 338) that's central to BPS practice. When our understanding of medical science excludes "information that is only accessible through the medium of human exchange" (Engel 1992, 338), he insists, we have misapplied the seventeenth-century paradigm in a way that compromises the goals of medical science.

This material is very helpful when it comes to the order of explanation between medical science and medical humanism. It's not that the BPSM advances a humanistic vision of patient as person, and then insists that medical science should adapt to humanism. On the contrary, Engel suggests that "appeals to humanism" are "ephemeral and insubstantial (…) when not based on rational principles" (Engel 1977, 135). We begin with

conceptual foundations, in other words, at the point where we clarify the scope and methods of medicine as a science, then this scientific vision forces us toward humanism (O'Leary 2021). Good medical science recognizes the relevance of biological, psychological and social factors, then it gathers data about those factors through a scientific approach to the clinical interview. That approach best succeeds when it humanizes patient and doctor, and in this sense, good science actually demands good ethics.

To my mind, this is Engel at his best, and all of this richness dissolves when we imagine that BPS practice could be a matter of simply applying psychosocial research in the clinic. Unfortunately, Engel's account of the clinical interview still leaves us entirely unclear about how to distinguish between biological, psychological and social explanations in the diagnostic process. Bolton and Gillett actually frame the question perfectly in Chapter 4:

> While disease is contextualised in the person as a whole, the immediate question is where the dysfunctional process is located: which system within the whole is dysfunctional, causing problems for the whole? (Bolton and Gillett 2019, 256)

Second, discourse about the BPSM, including Bolton and Gillette's, often fails to recognize how the lack of clinical guidance poses a threat to patient safety. When the model opens the door to psychosocial diagnosis for bodily symptoms in everyday practice, clearly it opens the door to a new and threating form of diagnostic error.

Diagnostic clarity is not the norm in medicine, surprising as that may be, at least not in outpatient care. In fact, as the UK's National Health Service understands things, "on average, 52% of patients accessing outpatient services have medically unexplained symptoms" (Joint Commissioning Panel for Mental Health 2017, 6-7). And while medical research and education are intensely focused on diagnosis, and treatment implied by diagnosis, they are essentially silent when it comes to developing directives for managing this very sizeable portion of cases.

Bolton and Gillett trust that "medical and clinical psychological textbooks" contain "scientific details" (Bolton and Gillett 2019, 119) that tell clinicians how to safely manage cases where biomedical and psychosocial explanations both remain possible, but that faith is wholly unfounded. Since the advent of the BPSM, recommendations for managing these cases have not been based on medical science at all, and they have not been evaluated by medical researchers for safety or reliability. Instead, practice in this area has been guided by research in psychiatry, specifically, research

produced and reviewed within the small subdiscipine of psychiatry known as psychosomatic medicine (or sometimes "consultation-liaison psychiatry").

Third, the need for clinical guidance has been met in psychosomatic medicine not by safety-tested science, but by wrangling about dualism. What makes the clinical problem so pressing, in other words, is that it combines in disastrous ways with the problem of ontological incoherence. In 1984, Schwab explained, for example, that according to "the established principles of psychosomatic medicine", in the great many cases where diagnosis remains elusive, clinicians should avoid "viewing the patient dichotomously as being 'organic or functional'" (Schwab 1985, 584). Instead of seeking clarity about the presence of disease, that is to say, a good BPS clinician will "conceptualize the patient as a total person, a psychobiological unit" (Schwab 1985, 584).

More recently, Creed and colleagues clarify the importance of avoiding "dualistic thinking" where we "regard symptoms as either organic or nonorganic/psychological". Instead, the BPS clinician should manage unexplained symptoms with deliberate diagnostic vagueness, making sure never to "force these disorders into either a 'mental' or 'physical' classification" (Creed et al. 2010, 5).

It is certainly possible for philosophical ideas to play a useful role in the challenge of distinguishing conditions with primarily biological causes from those with primarily psychosocial causes. Indeed, it's hard to see how we can understand that question without philosophical ideas about mind and body. Philosophy can be productive for medicine, though, only to the extent that it's supported with sound reasoning that's continuous with, and consistent with, science. In the borderlands between medicine and psychiatry, however, the BPSM's ontological confusion reaches its most incoherent pitch. Here Engel's defining demand to extend medicine's focus beyond body has somehow become a demand to equate mind with body at all times. The recommendation to see both mind and body as vital contributors to health has become a demand never to engage in practices that distinguish one from the other.

Even if we could defend these ideas in their own right, we cannot possibly defend them as consistent with the defining ideas of holism. More importantly, we cannot defend them as consistent with even the lowest standards for safety in medical science. By definition, cases of diagnostic uncertainty are cases where the possibility of biological disease remains, so these are cases where a recommendation to avoid biological clarity requires an extraordinarily high bar of scientific evidence. What it needs is

a consistent standard for determining when the possibility of biological disease can reasonably be set aside, and biomedical research that rigorously evaluates the safety of that standard for the wide range of patients who suffer from undiagnosed symptoms. What it has is the boogeyman of "dualism", an imagined imperative, borne of Engel's own confusion, to avoid diagnostic practice that "separates mind and body" at all costs.

Though medicine's research review system would root out these recommendations, research in psychosomatic medicine is not reviewed in the medical system. While medical textbooks and practice standards defer to psychosomatic medicine when it comes to principles for practice with medically unexplained symptoms, the research that drives these principles circumvents the filtering process for medical science. This too is the result of ontological incoherence. Because the BPSM proposes that biological and psychosocial factors are both relevant for medical practice, but it fails to provide guidance on how to manage that distinction, we have imagined that we can hand off vital matters of biomedical safety—for a very substantial portion of outpatients—to research and review within a subdiscipline of psychiatry. That, quite clearly, is a scientific mistake.

It should not be surprising that in the area where BPS ontology is poised to play its most direct and substantial clinical role, right there in the mind-body borderlands, we find recommendations for practice that are demonstrably problematic. Deep conceptual confusion rarely leads to empirical success for any science, and medicine is no exception to that rule.

4. Conclusions: New Holism

Bolton and Gillett's book is probably the best we can do when it comes to propping up the BPSM as a model for medical science. In that sense it may be most instructive by example. On the basis of the model itself, even with considerable philosophical ingenuity, we cannot escape the BPSM's entrenched philosophical confusions, and we cannot avoid the dangerous ramifications of those confusions in everyday practice.

Fortunately, we can reject the BPSM without accepting the biomedical model. In fact, we can reject it even as we accept that biological, psychological and social factors each play an inextricable role in human health. To do so is just to put our collective foot down, to insist that as holists we can do better, that the inchoate bag of ideas put forth by George Engel is both wise and inadequate, both essential and utterly absurd.

When a holist rejects the BPSM she does not advance a version of medicine where the patient becomes, once again, a body, where autonomy yields, once again, to parentalism. On the contrary, as a holist she holds those ideas in such high regard that she demands a sound foundation for them, a conceptual depth and consistency that's worthy of the task at hand. This demand is entirely in keeping with Engel's vision, with his suggestion that "appeals to humanism" are "ephemeral and insubstantial (…) when not based on rational principles" (Engel 1977, 135). Because humanism matters, we cannot achieve it on the cheap. To understand its roots, and its necessity, medicine needs to get its philosophical house in order.

The defining idea of holism is that medicine makes no sense, not in its humanity and not in its science, without the reality of human experience. We pursue the practice of medicine, and indeed we recognize it as morally imperative, because disease causes terrible experiences, and ultimately the cessation of experience. This point is so deeply obvious to those in the medical professions that it's a struggle even to imagine what it would mean for philosophers to question it, and to reject it, as they often do. It is helpful to note, too, that the reality of experience was no less obvious in medicine before Engel than it has been since. Regardless of the BMM's commitment to objective scientific methods, and regardless of its consensus that the realm of experience lies outside the scope of medicine, the medical profession has never denied, or even imagined denying, the reality of experience. It has always pursued medicine for the purpose of improving and protecting experience. It has always accepted facts of first-person experience as medicine's motivating data (O'Leary 2021).

In this sense, Engel's holistic vision was more a confession than a revelation. Without metaphysical specifics, it simply and broadly pointed out that human beings are experiencing beings, and that somehow, maintaining medicine's scientific commitment, we must recognize that in order for medicine to succeed. In effect, holism set out to position medicine's foundation somewhere within the framework of philosophy of mind, but with the BPSM that effort could not have been a more colossal failure. Not only has the BPSM failed to clarify medicine's philosophical position on mind and body. It has created, and in fact entrenched, a compendium of pseudo-philosophical jargon so incoherent as to make medical holism anathema to philosophy.

Holism should have inspired a conjoining of medicine with philosophy, a unified effort to understand experience in the context of medical science, and to apply that understanding to improve clinical practice. Instead, the language of the BPSM so distorted medicine's mind-body position that we now find ourselves demanding and rejecting dualism in the same breath—

not now and then, but as a defining feature of medicine's conceptual dogma (O'Leary 2020).

If we let go of the jumble of platitudes that is the BPSM—the equivocation on dualism, the unsupported prohibition on "separation", the imperative to "integrate" as if we have the power to change how mind and body are related—we can begin to fix this problem. We can accept medical holism as the metaphysical open door that it is, just a willingness to recognize the reality of experience, and the sense in which that reality forces medicine to address biological, psychological and social aspects of health. And we can finally characterize that perspective in accurate philosophical terms: as acceptance of consciousness in the context of medical science.[2]

This will not entirely resolve the question of medicine's position on dualism, and it will not explain how subjective experience can play a central role in objective medical science, but it will position medicine in the territory of nonreductive physicalism and property dualism, and that will make it possible to address medicine's basic ontological questions in a serious way. More than that, regardless of our answers to those questions, medical practice can readily be improved purely through recognition that a holist does distinguish conscious states from the brain states (or body states) with which they're correlated. This clarity makes it possible to develop practice recommendations for unexplained symptoms that are based on medical science rather than unsupported dogma about avoiding separation of mind and body.

In truth, we work with a placeholder in all fields where a sound philosophico-scientific picture of consciousness should be, and in this sense perhaps medicine can make an invaluable contribution. As an effort to improve and protect embodied experience through science, medicine is the mind-body problem writ large, with stakes that make the difference between wellness and suffering, health and disease, life and death for real persons. In a sense, medicine is the conscience of consciousness studies—or at least it would be if it took part. We are the applied science that keeps it real, the science that absolutely cannot do without experience as experience, the science where misunderstanding of mind and body will play out as real human suffering in the real world.

Bolton and Gillett are entirely right that "Engel's proposal of the biopsychosocial model was audacious" (Bolton and Gillett 2019, 89).

[2] By 'consciousness' I mean, specifically, phenomenal consciousness, following Block: "Phenomenal consciousness is experience; the phenomenally conscious aspect of a state is what it is like to be in that state. The mark of access-consciousness, by contrast, is availability for use in reasoning and rationally guiding speech and action" (Block 1995, 228).

What's audacious about it, though, is easy to miss. We take the reality of experience for granted in the context of medicine, and we take the possibility of medical science for granted, as well we should. What we should learn from Engel, most audaciously and most profoundly, is that we have work to do in sorting out how those truths fit together.

REFERENCES

Block, Ned. 1995. 'On a Confusion about a Function of Consciousness'. *Behavioral and Brain Sciences,* 18 (2), 227-87. https://doi.org/10.1017/S0140525X00038188.

Bolton, Derek and Grant Gillet. 2019. *The Biopsychosocial Model of Health and Disease New Philosophical and Scientific Developments*. Cham, Switzerland: Palgrave. https://doi.org/10.1007/978-3-030-11899-0.

Borrell-Carrió, Francesc, Anthony L. Suchman, and Ronald M. Epstein. 2004. 'The Biopsychosocial Model 25 Years Later: Principles, Practice, and Scientific Inquiry'. *Annals of Family Medicine* 2 (6): 576-82. https://doi.org/10.1370/afm.245.

Chalmers, David. 1995. 'Facing up to the Problem of Consciousness'. *Journal of Consciousness Studies* 2 (3): 200-219.

Creed, Francis, Elspeth Guthrie, Per Fink, Peter Henningsen, Winfried Rief, Michael Sharpe, and Peter White. 2010. 'Is there a Better Term than "Medically Unexplained Symptoms"?' *Journal of Psychosomatic Research* 68 (1): 5-8. https://doi.org/10.1016/j.jpsychores.2009.09.004.

Engel, George. 1977. 'The Need for a New Medical Model'. *Science*, 196 (4286): 129-36. https://doi.org/10.1126/science.

Engel, George. 1992. 'How Much Longer Must Medicine's Science be Bound by a Seventeenth Century World View?' *Family Systems Medicine* 10 (3): 333-45. http://dx.doi.org/10.1037/h0089296.

Foss, Laurence. 2002. *The End of Modern Medicine: Biomedicine Under a Microscope*. Albany: SUNY University Press.

Ghaemi, S. Nassir. 2009. 'The Rise and Fall of The Biopsychosocial Model'. *British Journal of Psychiatry* 195, 3-4. https://doi.org/10.1192/bjp.bp.109.063859.

Ghaemi, S. Nassir. 2010. *The Rise and Fall of the Biopsychosocial Model: Reconciling Art and Science in Psychiatry.* Baltimore: Johns Hopkins University Press. https://doi.org/ 10.1353/book.3501.

Joint Commissioning Panel for Mental Health. 2017. 'Guidance for Commissioners of Services for People with Medically Unexplained Symptoms'. https://www.jcpmh.info/wp-content/uploads/jcpmh-mus-guide.pdf.

Marcum, James A. 2008. 'Reflections on Humanizing Biomedicine'. *Perspectives in Biology and Medicine* 51 (3): 392-405. https://doi.org/10.1353/pbm.0.0023.

Miresco, Mark J., and Laurence J. Kirmayer. 2006. 'The Persistence of Mind-Brain Dualism in Psychiatric Reasoning about Clinical Scenarios'. *American Journal of Psychiatry* 163 (5): 913-18. https://doi.org/10.1176/ajp.2006.163.5.913.

O'Leary, Diane. 2020. 'Medicine's Metaphysical Morass: How Confusion about Dualism Threatens Public Health'. *Synthese*. https://doi.org/10.1007/s11229-020-02869-9.

O'Leary, Diane. 2021. 'The Value of Consciousness in Medicine'. In *Oxford Studies in Philosophy of Mind*, *Volume 1,* edited by Uriah Kriegel, 65-85. Oxford: Oxford University Press.

Schneider, Susan. 2012. 'Why Property Dualists Must Reject Substance Physicalism'. *Philosophical Studies* 157 (1): 61-76.

Schwab, John J. 1985. 'Psychosomatic Medicine: Its Past and Present'. *Psychosomatics* 26 (7): 583-85, 588-89, 592-93. https://doi.org/10.1016/S0033-3182(85)72821-6.

EuJAP | Vol. 17 | No. 2 | 2021
UDC: 122:159.95/.97
https://doi.org/10.31820/ejap.17.2.6

CAUSATION AND CAUSAL SELECTION IN THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE

Hane Htut Maung[1]

[1] University of Manchester

Original scientific article – Received: 27/01/2021 Accepted: 05/05/2021

ABSTRACT

In The Biopsychosocial Model of Health and Disease, Derek Bolton and Grant Gillett argue that a defensible updated version of the biopsychosocial model requires a metaphysically adequate account of disease causation that can accommodate biological, psychological, and social factors. This present paper offers a philosophical critique of their account of biopsychosocial causation. I argue that their account relies on claims about the normativity and the semantic content of biological information that are metaphysically contentious. Moreover, I suggest that these claims are unnecessary for a defence of biopsychosocial causation, as the roles of multiple and diverse factors in disease causation can be readily accommodated by a more widely accepted and less metaphysically contentious account of causation. I then raise the more general concern that they are misdiagnosing the problem with the traditional version of the biopsychosocial model. The challenge when developing an explanatorily valuable version of the biopsychosocial model, I argue, is not so much providing an adequate account of biopsychosocial causation, but providing an adequate account of causal selection. Finally, I consider how this problem may be solved to arrive at a more explanatorily valuable and clinically useful version of the biopsychosocial model.

Keywords: *Derek Bolton; Grant Gillett; biopsychosocial model; causation; causal selection*

Correspondence: hane.maung@manchester.ac.uk

1. Introduction

The biopsychosocial model, initially developed by George Engel (1977), is perhaps the most widely accepted model of health and disease in contemporary medicine. As the name suggests, the model emphasises the importance of considering biological, psychological, and social dimensions of health and disease in clinical practice. In recent years, however, the model has recently been criticised for being too vague to have any explanatory value or predictive power. The psychiatrist Nassir Ghaemi, for example, has suggested that the biopsychosocial model is not a scientific model, but is little more than "a slogan whose ultimate basis was eclecticism (…) meant to free practitioners to do what they pleased" (Ghaemi, 2010, p. 213).

Responding to this criticism, Derek Bolton and Grant Gillett aim to develop a defensible version of the biopsychosocial model that can support the theory and practice of contemporary medicine. In *The Biopsychosocial Model of Health and Disease* (2019), they propose that an appropriately updated version of the model can provide a philosophical framework which facilitates the understanding of disease causation. Given the increasing evidence that psychological and social factors have important roles in disease causation, they argue that physicalistic reductionism is false and that some version of the biopsychosocial model is required in medicine. However, a problem with the traditional version of the biopsychosocial model is that it does not tell us how these biological, psychological, and social factors interact causally. Accordingly, they suggest that a suitably updated version of the model must include a metaphysically adequate account of biopsychosocial causation that can accommodate the roles of these multiple and diverse factors.

In this paper, I offer a philosophical critique of the analysis of biopsychosocial causation provided by Bolton and Gillett. While I agree with them that physicalistic reductionism is untenable and that some version of the biopsychosocial model is warranted, I argue that their causal approach to defending the model is problematic. In §2, I briefly lay out the account of biopsychosocial causation provided by Bolton and Gillett. In §3, I show that their account relies on claims about the normativity and the semantic content of biological information that are metaphysically contentious. Moreover, I suggest that these claims are unnecessary for a defence of biopsychosocial causation, as the roles of multiple and diverse factors in disease causation can be readily accommodated by a more widely accepted and less metaphysically contentious account, namely James Woodward's (2004) interventionist theory of causation. In §4, I raise a more general worry, which is that Bolton and Gillett are misdiagnosing the

problem with the traditional version of the biopsychosocial model. The key challenge when developing an explanatorily valuable version of the biopsychosocial model, I suggest, is not so much providing a metaphysically adequate account of causation, but providing an epistemically useful account of causal selection. That is to say, the vagueness of the biopsychosocial model is related to its inability to tell us which causal factors, out of the vast network of biological, psychological, and social factors, are explanatorily significant. Finally, I consider how this problem may be solved to arrive at a more explanatorily valuable and clinically useful version of the biopsychosocial model.

2. An Account of Biopsychosocial Causation

The traditional version of the biopsychosocial model presented by Engel (1977) arose in response to the prevailing model in medicine at the time, which was the biomedical model of health and disease. This is characterised as follows:

> It assumes disease to be fully accounted for by deviations from the norm of measurable biological (somatic) variables. It leaves no room within its framework for the social, psychological, and behavioral dimensions of illness. The biomedical model not only requires that disease be dealt with as an entity independent of social behavior, it also demands that behavioral aberrations be explained on the basis of disordered somatic (biochemical or neurophysiological) processes. (Engel 1977, 130)

A key feature of the biomedical model, then, is physicalistic reductionism, or the assumption that disease can be reductively explained at the lowest biological level, which may be biochemical or neurophysiological. Psychological and social factors are either excluded from the explanation or assumed to be reducible to processes at the biological level.

While the biomedical model is supported by advances in biomedical science, Engel argues that it has serious limitations that make it insufficient as a general model for medicine. These include its neglect of the patient's account of the illness, its inability to consider how social circumstances influence the presentations and meanings of health and disease, and its failure to acknowledge the roles of psychological and social factors in disease causation. In their book, Bolton and Gillett spend considerable time on the last of these, citing the accumulating evidence that psychological and social factors have causal roles in health and disease. They list a wide range of conditions that are influenced by psychological and social factors:

> For example: breast cancer (…) atopic disease, generally, including for asthma; HIV and musculoskeletal disorders. In addition, psychosocial factors have been implicated in outcomes of surgical procedures, for example, chronic pain; lumbar and spinal surgery; liver transplant (…) and coronary artery bypass (…) In addition, there is evidence for psychosocial factors in wound healing, and extent of fatigue after traumatic brain injury. Psychosocial factors have also been implicated in responses to other interventions for medical conditions, such as inpatient rehabilitation for stroke patients (…) and effects of hospitalisation on older patients. (Bolton and Gillett 2019, 11–12)

The above is supported by the extensive epidemiological research of Michael Marmot (2005), who demonstrated robust correlations between social statuses and the incidences of a wide range of medical conditions. Hence, just as the biomedical model is of interest because of the advances in biomedical science, the biopsychosocial model is supported by advances in psychology, epidemiology, and social science.

In the present day, the contributions of psychological and social factors are especially apparent in the increasing rates of mental health problems in young people. Bolton explores some of these factors in a recent paper coauthored with the psychiatrist Dinesh Bhugra (Bolton and Bhugra, 2020). They argue that changes in society over the past few decades have contributed to worsening mental health problems among children, adolescents, and young adults. For example, due to the development of social media and the public profile of populism, political conflicts between conservatives and liberals have become more visible and pervasive in ways that have eroded the shared norms of rationality in political discourse and have resulted in the loss of social cohesion. Moreover, due to government austerity, neoliberal financialisation, and economic downturn, intergenerational wealth inequalities have increased, with young adults from the millennial generation having less stable accommodation, less career certainty, and less financial security than older adults from the baby boomer generation. The negative mental health effects of these economic and political factors are corroborated by epidemiological data showing that invoking government austerity during an economic recession increases the population suicide rate, while investing in social welfare during an economic recession does not have this outcome (Stuckler and Basu, 2013). Finally, younger generations are also affected by serious concerns regarding anthropogenic climate change and the inadequate geopolitical response to the environmental crisis.

Given that neither the genetic nor the neurobiological characteristics of people have changed significantly over the past few decades, the biomedical model appears inadequate to account for these increasing rates of mental health problems in young people. Rather, Bolton and Bhugra (2020) argue that a broad biopsychosocial approach is required to account for the contributions of the aforementioned changes in society to these worsening mental health problems. Accordingly, in their book, Bolton and Gillett (2019) develop a metaphysical account of causation that avoids the physicalistic reductionism of the biomedical model and accommodates the roles of biological, psychological, and social factors in disease causation.

Against physicalistic reduction, Bolton and Gillett argue that explanations in biology are irreducible to explanations in chemistry and physics. Following the work of Erwin Schrödinger (1944), they suggest that biological systems are characterised by their abilities to extract energy from the environment and resist local increases in entropy, thus allowing them to maintain stable forms, develop in ordered ways, and reproduce. According to Bolton and Gillett, biological systems can do this because they use information transfer to control energy transfer. They write:

> Physical and chemical processes involve energy transfers covered by mathematical energy equations, but in biological organisms the physical and chemical processes not only happen, but can only happen in the right place at the right time in the right degree, if there are mechanisms that control and regulate them in a way appropriate to bringing about a particular function. (Bolton and Gillett 2019, 48)

The informational nature of biological causation, Bolton and Gillett argue, is irreducible to physical explanation, because it involves semantic content. The dynamics of this semantic content follow regularities that are not captured by the lawlike regularities of physics and chemistry. Bolton and Gillett continue:

> Another way of making this point is that the energy transfer involved in information transfer is irrelevant to the information transfer. The flow of information depends on regularities, but these regularities are not determined by the energy equations of physics and chemistry, rather they must rely on other properties of materiality. The concept required at this point is expressed by such terms as *structure*, *form*, *shape* or *syntax* (to borrow from logic)—that codes information. (Bolton and Gillett 2019, 49)

For example, sequences of nucleotides on genes encode information that is used by intracellular components to construct proteins, patterns of action potentials in neurons encode information that influence how neurotransmitters are secreted, and ligands encode information in virtue of their selective interactions with receptors.

Bolton and Gillett go on to argue that the semantic content of biological information makes biological causation normative and teleological. That is to say, there are "right" and "wrong" ways for the semantic content to be decoded, which pertain respectively to whether or not they are conducive to the biological systems fulfilling their goals or functions. Such normativity, Bolton and Gillett suggest, makes causation in biology different from causation in physics. While causation in biology is characterised by the capacity for error, causation in physics is purported to follow laws and equations that cannot be violated. They write:

> The general conceptual point at issue here is that regulation and control mechanisms keep things going *right rather than wrong*. Such normativity is not present in the energy equations of physics and chemistry, which always apply and never fail. It arises in biology for the first time, marking a fundamental departure of biology from physical and chemical processes alone. The normativity is implied in all of the key systems theoretic concepts such as regulation, control and information. It derives from the point that biological systems function towards ends, and function well and badly accordingly as they do or do not attain them. (Bolton and Gillett 2019, 51)

For example, at the genetic level, the sequences of nucleotides are usually conserved during genetic replication, but mutations occasionally occur due to "replication errors", some of which can have harmful effects for the organisms. At the molecular level, immunoreceptors usually bind selectively with particular foreign ligands, but occasionally they react with antigens from hosts due to "molecular mimicry", which can be associated with autoimmune reactions. At the organismal level, a behaviour, such as feeding, is usually adaptive insofar as it contributes to the survival and reproduction of the organism, but occasionally may be maladaptive, such as when it leads to the ingestion of a toxin.

Informational content and normativity are also characteristics of psychological and social processes. For example, perception can be deemed accurate or inaccurate according to perceptual norms, belief can be deemed rational or irrational according to epistemic norms, speech may be deemed correct or incorrect according to linguistic norms, and

behaviour can be deemed permissible or impermissible according to moral, legal, and social norms. Bolton and Gillett suggest that these interact with the informational content and normativity of biological processes through embodied agency. They draw on a recent development in the philosophy of mind, which Albert Newen, Leon De Bruin, and Shaun Gallagher call 4E cognition (Newen et al. 2018). This proposes that cognition has the four following features:

1. 'Embodied' (in the body)
2. 'Embedded' (in the environment; in causal loops with it)
3. 'Enactive' (Acting in and manipulating the environment, directly, not via a representation or model; the environment offers affordances, or opportunities, for action and manipulation)
4. 'Extended' (Extended to the body and environment, including devices used for cognitive functioning). (Bolton and Gillett 2019, 78)

Psychological agency, according to Bolton and Gillett, is embodied in the biological body and, in virtue of the informational transfer that occurs in the biological body, is an active causal power whose influence extends into the social environment. Accordingly, normative processes at biological, psychological, and social levels can interact with one another causally via the regulatory flow of information.

To bring this all together, let us see how it might apply to the aforementioned increasing rates of mental health problems among young people (Bolton and Bhugra, 2020). Recent social and political changes, including the shared norms of rationality in political discourse being undermined, increasing intergenerational wealth inequalities, and escalating concerns about anthropogenic climate change, lead to adverse social conditions. These have downward regulatory effects that restrict psychological agency, constrain how biological resources are distributed, and disrupt the usual flow of information in the biological system. In turn, the alteration in the informational transfer in the biological system further affects psychological agency and disrupts how the person interacts with the social environment, manifesting in mental ill health.

Here, the biological, psychological, and social processes are integrated, with information transfer being the common currency in the causal interactions across these three domains. This information transfer has a normative dimension that is irreducible to the sort of causal explanation that features in physics. And so, the account of biopsychosocial causation developed by Bolton and Gillett (2019) accommodates the roles of

multiple and diverse factors in disease causation while avoiding the physicalistic reductionism of the biomedical model. However, their account relies on claims about the normativity and semantic content of biological information that are metaphysically contentious. In the following section, I examine some of the problems with these claims and show that they are unnecessary for an adequate account of biopsychosocial causation.

3. Critical Discussion

Bolton and Gillett are indeed correct that informational content and normativity are properties of the psychological and social domains respectively. Psychological agency is marked by intentionality and meaning, which are embedded in the wider social context and appear to be irreducible to the regularities studied in physics. The social environment is marked by our values, norms, and conventions, which regulate our behavioural affordances, interpersonal interactions, and communicative practices. Hence, informational content and normativity in the psychological and social domains have their sources in our intentions, values, interests, and judgements at the interpersonal level. However, claiming that normativity and informational content are properties of the biological domain at the subpersonal level is more problematic. Of course, Bolton and Gillett are correct that we often use normative and informational notions, such as function, dysfunction, sense, and error, in biological theorising. The problem, though, is that these normative and informational notions may be features that we project onto biological processes, rather than intrinsic properties of the biological processes themselves. That is to say, we derive notions from our understandings of the genuine normativity and informational content of the social and psychological domains, and then we use these notions as instrumental metaphors to organise our theoretical thinking about biological processes.

The above presents challenge to the account of biopsychosocial causation presented by Bolton and Gillett for the following reason. As noted above, information transfer is supposed to be the common currency in the causal interactions across biological, psychological, and social domains. However, if normativity and informational content are not genuine properties of biological causation but are merely instrumental metaphors that we use to organise our theoretical thinking about biological processes, then such information transfer cannot comprise the common currency that is conserved across the three domains in biopsychosocial causation. Causation in the psychological and social domains may involve genuine

normative and informational properties, but it is doubtful whether these properties can actually be said to be conserved at the biological level.

My contention that normative and informational notions in biology are instrumental metaphors can be illustrated in two ways. First, I consider how mechanical laws and explanations in physics might be rephrased in teleological and normative terms. This challenges the claim by Bolton and Gillett that normativity is what makes causation in biology different from causation in physics. Second, I consider how explanations in biology that invoke normative and informational notions might be rephrased in terms that are more descriptive. This challenges the claim that normativity and informational content are intrinsic properties of the biological processes themselves.

With respect to causation in physics, recall that Bolton and Gillett claim that this follows laws and equations that cannot be violated, in contrast with causation in biology which they claim is capable of error. However, the regularities in physics may not be as faultless as Bolton and Gillett suggest. Suppose, for example, that a trolley with a known mass is attached to a hanging stone of a known weight via a pulley and the acceleration of the trolley is measured. The theoretical law in this case is $F = m \times a$, where F is the total pulling force of the hanging weight, m is the mass of the trolley, and a is the acceleration of the trolley. Now, if the experiment is repeated under a variety of background conditions, a may turn out not to be the same in each instance despite F and m being kept constant. That is to say, the observations may deviate from what is predicted by $F = m \times a$ in different ways.

As noted by Imre Lakatos (1974), when this happens, we tend to invoke auxiliary hypotheses which introduce other variables, in order to conserve $F = m \times a$. For example, we may try to explain the variability in a across the different experimental conditions by considering possible confounding factors, including variations in the energy lost through friction, air resistance, and elasticity of the cord attaching the trolley to the weight. However, our hypotheses based on these confounding factors may not be able to yield quantities that are sufficiently exact to conserve $F = m \times a$. Indeed, as Nancy Cartwright (1983) points out, solving the derived equations to see whether or not they fit with our observations may be mathematically intractable. For example, if we try to derive the energy lost through friction from the mechanical and thermodynamic properties of the trolley and the surface, and then try to predict how this would affect the movement of the trolley at different moments in its trajectory, we may only yield rough approximations. Hence, far from being faultless, the

regularities in physics are associated with various deviations for which we may not be able to account mathematically.

This capacity for error in physics raises the possibility of rephrasing mechanical laws and explanations in teleological and normative terms, akin to explanations in biology. To take another example, consider the law that a system comprising two objects in contact with each other will proceed toward thermal equilibrium. This can be rephrased as a teleological and normative claim, whereby proceeding towards the "goal" of thermal equilibrium is what the system "should" do. However, in actuality, systems tend not to be closed, and so may involve thermal disequilibria that deviate from this law. These could be interpreted as cases where contingent circumstances result in the systems "failing" to proceed as they "should", analogous to dysfunctions in biological systems. An objection might be to say that while there can be localised thermal disequilibria, the universe as a whole is proceeding toward thermal equilibrium, which will eventually result in these localised thermal disequilibria being dissipated. In response, though, an analogous claim could be made regarding dysfunctions in biological systems. That is to say, while there can be localised dysfunctions that compromise the survival and reproductive prospects of organisms, but it could be claimed that the frequencies of these dysfunctions will eventually diminish through the process of natural selection.

Of course, these teleological and normative notions are not intended to be literal. That is to say, they involve no ontological commitment to the claim that systems in physics actually have "goals". Rather, they are instrumental metaphors that are derived from the teleological and normative notions we use in the psychological and social domains, which concern our intentions, values, interests, and judgements. Nonetheless, the possibility of rephrasing regularities in physics in teleological and normative terms suggests that they may not necessarily be so different from regularities in biology. It gives us grounds to consider whether the teleological and normative notions in biological explanations are also instrumental metaphors, rather than being representations of actual properties of biological processes. To be clear, this is not to say that biological explanation can be reduced to physical explanation. I agree with Bolton and Gillett that the complex causal processes in biology are not straightforwardly reducible to the mechanical laws and explanations in physics. Rather, it is to say that the difference between the domains of biology and physics cannot be captured by the presence or absence of normativity. This can be further demonstrated by examining how teleological and normative explanations in biology can be rephrased in terms that are more descriptive.

With respect to causation in biology, recall that Bolton and Gillett claim that this is characterised by informational content that can be decoded in "right" or "wrong" ways, which pertain respectively to whether or not they are conducive to the biological systems fulfilling their goals or functions. At the genetic level, they suggest that information is encoded in the sequences of nucleotides on chromosomes and, if decoded properly, contributes to the proper forms of the biological systems being maintained. Here, Bolton and Gillett seem to adhere to the modern evolutionary synthesis, which considers the genome to be a "blueprint" for the realisation of the phenotype (Plomin, 2018). A notable proponent of this view is Richard Dawkins, who suggests that the "information passes through bodies and affects them, but it is not affected by them on its way through" (Dawkins 1995, 4).

However, recent developments in the philosophy of biology have undermined the modern evolutionary synthesis. An important contribution is a theoretical framework, put forward by Susan Oyama, Paul Griffiths, and Russell Gray, called developmental systems theory (Griffiths and Gray, 1994; Oyama, 2000). Developmental systems theory emphasises that the genome is just one among many dynamic resources that interact to produce a phenotypic outcome, including epigenetic modifications, transcription factors, intracellular reactions, physiological processes, nutritional resources, environmental conditions, social interactions, and cultural contexts. That is to say, the phenotype is not the inevitable realisation of a genetic "blueprint", but is the contingent outcome of complex and dynamic interactions between multiple resources, some of which may also be inherited across generations. Variations in these resources can result in variations in the phenotypic outcomes. Accordingly, Griffiths and Gray (1994) argue that the genome cannot be considered to be a unique bearer of developmental information. Given that the particular causal role of the genome is contingent on the state of the rest of the developmental system, it makes just as much sense to say that the rest of the developmental system encodes information that is "read" by the genome as it does to say that the genome encodes information that is "read" by the rest of the developmental system. Informational content, then, is not an intrinsic property of biological causation, but is an instrumental metaphor whose application depends on what part of the developmental system we decide to hold fixed. As Oyama notes, information is just "a way of talking about certain interactions rather than their cause or a prescription for them" (Oyama 2000, 197).

The contingency and multifactoriality of development challenge the view that teleology and normativity are inherent in biological causation. Instead of there being "right" and "wrong" ways to decode a sequence of

nucleotides, there are just different causal outcomes that can result from different combinations of interacting factors. For example, at the level of genetic replication, we can think of "replication errors" not as literal mistakes, but as different causal outcomes of the interactions between nucleotides and polymerases due to the influences of external factors and variations in intracellular conditions, much like how the deviations from $F = m \times a$ in the trolley experiment are different outcomes that result from differences in the experimental conditions. At the level of phenotypic development, we can think of different outcomes not as expressions of the genotype gone "right" and gone "wrong", but as different contingent forms that result from different developmental conditions. For example, genetically indistinguishable specimens from the fish species *Salmo trutta* can develop into the small freshwater brown trout or into the large saltwater sea trout, depending on the ecological conditions in their early developmental stages (Charles *et al.* 2005). These forms are morphologically and behaviourally different, but are both capable of thriving and reproducing. Neither form represents the "right" way to realise the *Salmo trutta* genome, but rather both are different causal outcomes that result from different combinations of developmental resources.

At this point, it might be contended that it is possible to discern "right" and "wrong" ways for biological systems to develop by considering whether or not parts of these biological systems are performing their functions. For example, a "replication error" that occurs during genetic replication may be considered to be an instance of the system going "wrong" if it compromises the ability of the resulting cell to function properly. However, this would be to concede that teleology and normativity are instrumental metaphors we project onto biological processes rather than properties of the processes themselves. As Matthew Ratcliffe notes, functions are not found out there in the world, but are contributions to goals "which are themselves instrumentally assigned" (Ratcliffe 2000, 124). That is to say, we instrumentally assign goals to systems and then assign functions relative to those goals. Parts of the systems are deemed to be functional if their effects are conducive to achieving these assigned goals in appropriate ways and are deemed to be dysfunctional if they are failing to produce these effects.

Usually, in biological enquiry, the assigned goal is survival of the biological system. Assigning this goal provides a focus which facilitates questions such as "what is it that x does to contribute to survival?" and "how did it come to do this?" (Ratcliffe 2000, 129). The former question is typically associated with Robert Cummins' (1975) functional analysis of the causal roles of parts of systems, while the latter question is typically

associated with Ruth Millikan's (1984) aetiological account of function based on the adaptive benefits of the effects of the parts in the evolutionary histories of organisms. While these questions are arrived at through the prior instrumental assignment of a goal, the answers can be expressed in causal and historical terms that do not invoke teleology. For example, we may assign an organism's retina the function of light transduction, because light transduction is the effect of the retina that contributes to the assigned goal of survival. From here, we might go on to explain how light transduction increases the likelihood of survival by influencing the organism's interaction with the environment. We might also go on to explain how the retina came to transduce light by giving a causal account of how past organisms with cells that transduced light had higher chances of producing offspring than past organisms without these cells, which resulted in the evolutionary transmission of the capacity for light transduction to the present organism. The assignment of function provides a focus, but the subsequent explanations are causal and historical explanations that do not themselves invoke a future goal or desired outcome. The normative notions of function and dysfunction, then, are not properties of the causal processes themselves, but are judgements we make relative to the goals we assign.

To further illustrate the instrumentality of function ascription in biology, consider the example of an alteration in an oncogene caused by exposure to an environmental carcinogen. The altered oncogene causally contributes to the accelerated proliferation of malignant tissue containing the altered genotype, which results in tumour progression. Usually, we would consider the alteration in the oncogene to be a dysfunction relative to the assigned goal of survival of the organism. However, it is at least theoretically possible to consider it to be properly functional if a different goal is assigned at a different level of analysis. For example, if we focus on the level of the tumour instead of the level of the organism, then we could claim that the function of the altered oncogene is the proliferation of malignant tissue, insofar as this is the effect of the altered oncogene that contributes to maintenance and progression of the tumour. Furthermore, this could be supported by the aetiological account of function, as the accelerated proliferation of malignant tissue is the effect of the altered oncogene that resulted in the abundance of the altered genotype in the developing tumour. Nonetheless, we tend not to consider the proliferation of malignant tissue to be the function of an altered oncogene, because we tend not to assign a goal at the level of the tumour. Rather, we tend to ascribe the goal of survival at the level of the organism and, accordingly, to consider the proliferation of malignant tissue to be a dysfunction relative to this goal. Hence, as Valerie Hardcastle notes, the assignment of function

is influenced by a value judgement about which level of analysis is "worthy of teleological language" (Hardcastle 2002, 149).

And so, teleology and normativity are not intrinsic properties of biological processes themselves, but are instrumental metaphors we project onto the biological processes. Biological systems are judged to go "right" or "wrong" relative to goals we assign to them. These normative notions and instrumental goals are derived from our understandings of genuine normativity and teleology in the psychological and social domains. For example, we consider survival of the organism, but not the progression of a tumour, to be a goal, partly because we judge surviving to be valuable and instrumental to our attaining our personal and collective aims and interests. As noted earlier, the informational and semantic notions that are employed in biological theorising are also derived from our understandings of information transfer and semantic content in the social and psychological domains.

The above poses a problem for the account of biopsychosocial causation presented by Bolton and Gillett (2019), because it suggests that normativity and information transfer cannot serve as the common currency in the causal interactions across these three domains. Information transfer and normativity are features of the psychological and social domains respectively, as these involve meanings, intentions, values, and interests. While we may invoke these notions in biological theorising, their uses are metaphorical and do not involve any ontological commitment to the claim that normativity and informational content are properties of the biological systems themselves. Hence, there is no good reason to suppose that the normative and informational notions we invoke in biological explanations refer to the same sorts of normativity and information transfer that feature in social and psychological explanations. There remains a disunity between the interpersonal level and the subpersonal level.

This brings us to the question of whether or not the above undermines the prospect of a philosophically defensible version of the biopsychosocial model. I argue that it does not. Recall that Bolton and Gillett present their account of biopsychosocial causation in order to accommodate the roles of multiple and diverse factors in disease causation while avoiding the physicalistic reductionism of the biomedical model. Accordingly, they suggest that biological, psychological, and social processes are normative processes that regulate one another through information transfer. However, there is no need for Bolton and Gillett to rely on such a metaphysically contested thesis in order to make sense of biopsychosocial causation. The fact that social factors causally influence biological outcomes is uncontroversial in contemporary healthcare and epidemiological research

has been able to demonstrate these causal relations without having to assume stronger metaphysical claims about biological processes.

Indeed, there is a more established philosophical account of causation that is more metaphysically neutral and can accommodate the roles of diverse factors. This is Woodward's (2004) interventionist theory of causation, which proposes the following:

> A necessary and sufficient condition for *X* to be a (type-level) direct cause of *Y* with respect to a variable set **V** is that there be a possible intervention on *X* that will change *Y* or the probability distribution of *Y* when one holds fixed at some value all other variables Z_i in **V**. (Woodward 2004, 59)

That is to say, causation is analysed as a probabilistic counterfactual dependence relation, wherein *X* is a cause of *Y* if and only if an intervention that changes *X* makes a difference to the probability of *Y* given appropriate background conditions. Importantly, no ontological restrictions are placed on what sorts of factors can be difference makers. Causal relations between factors can be established by using interventions to demonstrate probabilistic dependencies between the factors, regardless of the organisational levels to which these factors belong. Accordingly, the interventionist theory of causation can accommodate causal relations between factors across biological, psychological, and social domains.

Bolton and Gillett do cite Woodward's interventionist theory of causation in their book. Specifically, they suggest that the interventionist theory of causation is consistent with their claims about agency and causation, insofar as it "emphasises that our interests in causal connections and explanations are linked to our practical concerns of being able reliably to bring about changes" (Bolton and Gillett 2019, 83). The problem, however, is that accepting the interventionist theory of causation makes their metaphysical claims about the normativity and informational content of biological processes somewhat superfluous. As John Campbell (2016) notes, if we understand causal relations in terms of probabilistic dependencies between factors that can be analysed counterfactually, then we do not need to commit to such stronger metaphysical claims in order to make sense of how biological, psychological, and social factors can interact in disease causation. Of course, further scientific research may later yield hypotheses about the mechanisms involved in some, though maybe not all, of these causal relations, but such mechanistic details are not necessary to establish that the factors are causally related.

The interventionist theory of causation also rejects the physicalistic reductionism of the biomedical model. By understanding causal relations in terms of probabilistic dependencies between factors, psychological and social factors can be acknowledged as genuine causal factors that make differences to biological outcomes, while also accepting that these psychological and social factors may be irreducible to biological processes. For example, recall the various social, political, and economic factors that Bolton and Bhugra (2020) suggest to be contributors to the increasing rates of mental health problems among young people. We can understand these factors as being causal in virtue of how changes in them make differences to the health outcomes when other variables are held fixed. David Stuckler and Sanjay Basu (2013) demonstrate such a causal relation between government austerity and an increase in the population suicide rate by comparing this situation to contrastive scenarios where different policies are associated with different outcomes. Here, establishing such a causal relation requires neither any attempt to reduce government austerity to a different explanatory level, nor any ontological commitment to some deeper property that is conserved or transmitted throughout the causal process.

4. The Problem of Causal Selection

The discussion so far suggests that biopsychosocial causation does not have to be so metaphysically taxing. It is widely accepted that social factors can influence biological outcomes and the interventionist theory of causation allows us to make sense of this without having to commit to further ontological claims about the normativity or informational content of biological causation. This raises the question of whether Bolton and Gillett (2019) have misdiagnosed the problem with the traditional version of the biopsychosocial model.

As noted earlier, Ghaemi (2010) criticises the biopsychosocial model for being too vague and too eclectic to have any explanatory value. Such eclecticism, he suggests, was "meant to free practitioners to do what they pleased" (Ghaemi 2010, 213). However, the problem raised by this criticism is not that the biopsychosocial cannot make sense of how the three domains interact causally, but rather that it includes so many causal factors that it does not offer a precise explanation. Alex Broadbent raises a similar worry about the multifactorial model of disease, noting that "[b]are multifactorialism does nothing to encourage the move from a catalogue of causes to a general explanatory hypothesis" (Broadbent 2009, 307). That is to say, listing more causal factors and causal relations does not necessarily make a model more explanatory.

The challenge when developing a defensible version of the biopsychosocial model, then, is not so much providing an adequate account of biopsychosocial causation, but providing an adequate account of causal selection. As Broadbent (2009) notes, under the conventional philosophical view of causation, almost every event that is caused is the outcome of multiple causal factors. Nonetheless, we only consider some of these causal factors to be relevant in an explanation. For example, when we want an explanation of house fire, we consider the electrical fault and the building's cladding to be explanatorily relevant, but not the presence of oxygen in the atmosphere, even though the accident was also causally dependent on this. Likewise, given that the biopsychosocial model does not exclude any sorts of causal factors *a priori*, it is trivially true that every disease is caused by multiple biological, psychological, and social factors. However, this does not tell us which of these factors are relevant in an explanation of the disease.

To some extent, the question of which causal factors are explanatorily relevant is an empirical issue, as we might be able to demonstrate empirically that different cases instantiate different combinations of causal factors. However, it is also to a significant extent a superempirical issue, as we still need to judge which of the many causal factors instantiated by a given case are explanatorily relevant and which comprise the background conditions. For example, we can catalogue all of the causal factors that contribute to a person's type II diabetes mellitus, including insulin resistance, altered β-cell activity, learned eating behaviour, sedentary labour, economic inequality, and the structure of the food environment, but cataloguing these factors will not inform us which of these factors are deemed explanatory and which are deemed to be in the background, nor will it inform how we should approach the problem. By contrast, the biomedical model fails for dismissing psychological and social factors, but offers a more specific guide to explanation and intervention, insofar as it privileges the biological level as the proper level of analysis.

There are two possible ways in which we might enhance the explanatory power of the biopsychosocial model. The first potential approach is to supplement the biopsychosocial model with a conceptual criterion for selecting explanatory factors from background factors. For example, factors may be deemed more explanatory based on causal proximity, speed of response, or specificity of response (Ross 2018). However, the problem with this approach is that setting *a priori* constraints on what factors are privileged as explanatorily relevant would revert back to a form of reductionism that the biopsychosocial model is seeking to avoid. Indeed, the physicalistic reductionism of the biomedical model could be interpreted as its assumption of biological proximity as a conceptual

criterion for which factors are deemed explanatory. Also, a further problem with this approach is that it ignores the different contexts in which different factors might be deemed explanatorily relevant. In different settings, the most explanatorily relevant factors may not be the most proximal, the fastest, or the most specific factors. For example, in a public health context, poor sanitation may be considered a very explanatorily relevant cause of cholera, even though it is not the most proximal cause, the cause with the fastest action, or a cause that is specific to cholera.

This brings us to the second potential approach. This is to acknowledge that which causal factors are deemed explanatory and which are deemed to be in the background are dependent on contexts, values, and interests. As Peter Lipton (2004) notes, explanations are not *tout court*, but are relative to contrastive foils. For example, when we ask "why did the leaves turn yellow?", the relevant answer will differ depending on whether we are asking "why did the leaves turn yellow in November rather than in January?" or "why did the leaves turn yellow rather than blue?" (Lipton 2004, 33). This suggests that in order for the biopsychosocial model to be explanatorily useful, we have to be more explicit about our explanatory interests and more specific about the questions we ask. Instead of asking what causes a disease *tout court*, we can yield more precise causal explanations by considering which contrastive foils are appropriate in the contexts and by asking more specific questions relative to these contrastive foils.

As well as being informed by epistemic and pragmatic considerations, our explanatory interests are often informed by ethical and political considerations, especially in healthcare, where promoting people's welfare and alleviating their suffering are central values. For example, in their recent research on transgender mental health, Sav Zwickl and colleagues apply a psychosocial approach to examine the causal factors associated with suicidality among transgender and nonbinary adults (Zwickl *et al.* 2021). The context of this research pertains to the higher rates of suicidality and mental health problems among transgender and nonbinary people than among cisgender people, and so the explanatory interests guiding the research are appropriately informed by ethical and political considerations concerning health inequity, social injustice, and systemic discrimination. Guided by these explanatory interests, the researchers were able to discern causal factors for suicidality that disproportionately or specifically affect transgender and nonbinary people, including lack of access to gender affirming healthcare, institutional discrimination, and transphobic violence. These causal factors could have been missed had different explanatory interests guided the research, such as a more general emphasis on the aetiology of mental illness rather than a more specific emphasis on

the mental health disparities between transgender people and cisgender people.

The above suggests that the biopsychosocial model complements a form of explanatory pluralism in healthcare. Given that it places no *a priori* constraints on what domains can be causal, it allows for a range of contexts that may require different explanatory approaches. This is noted by Leen De Vreese and colleagues, who suggest that the question "why did person P develop lung cancer?" can allow for many relevance relations, including the following:

> (a) Why did person P, who smokes, develop lung cancer, while person P', who also smokes, did not?
>
> (b) Why did person P with behavior B develop lung cancer, while person P' with behavior B' did not?
>
> (c) Why did person P living in country C develop lung cancer, while person P' in country C' did not? (De Vreese *et al.* 2010, 375–376)

The different relevance relations warrant explanations that appeal to causal factors from different domains. Question (a) is about how a physiological difference between the two people results in smoking having different effects, and so calls for a physiological explanation that draws on biological factors. Question (b) is about the difference between the behaviours of the two people, and so calls for a behavioural explanation that draws on psychological factors. Question (c) is about the effects of the different environments of the two people, and so calls for an epidemiological explanation that draws on social factors.

In turn, the answers to these questions can inform preventative and therapeutic interventions across different healthcare disciplines. For example, the answer to (a) could inform targeted screening and oncological treatment, the answer to (b) could inform behavioural and cognitive interventions such as smoking cessation therapy and motivational counselling, and the answer to (c) could inform public health interventions such as smoking policies and clean air strategies. And so, if we are explicit about our explanatory interests and ask appropriately specific questions, the biopsychosocial model can support clinical interventions that target causal factors across multiple domains.

Of course, explanatory pluralism is not a new idea in the philosophy of medicine. For example, Kenneth Kendler (2005) and Sandra Mitchell

(2009) have endorsed pluralistic approaches to explaining mental disorders that consider causal factors at genetic, neurobiological, psychological, interpersonal, and cultural levels. However, while the form of explanatory pluralism endorsed by Kendler and Mitchell is an integrative pluralism that seeks to integrate the diverse causal factors at multiple levels into a single comprehensive model, the form of explanatory pluralism I am proposing does not require such integration. Rather, given the biopsychosocial model's wide interdisciplinary scope, it may sometimes be better complemented by a looser form of ineliminative pluralism akin to that suggested by Helen Longino (2013) for studying behaviour. That is to say, we may understand disease causation better by utilising multiple partial accounts than by attempting to assemble a more general model that incorporates all the causal factors. Different partial accounts may be relevant to different explanatory interests and may draw on different sets of causal factors. For example, in response to the aforementioned question "why did person P develop lung cancer?", whether we consider a predominantly physiological account, a predominantly behavioural account, or a predominantly epidemiological account to be appropriate will depend on the relevance relations in which we are interested (De Vreese *et al.* 2010). It may not be possible to integrate these accounts into a single comprehensive model that represents all of the causal relations between the different domains, but this does not compromise the clinical value of the biopsychosocial model.

5. Conclusion

Bolton and Gillett (2019) are correct that there is good reason to endorse the biopsychosocial model in contemporary healthcare. Given the substantial evidence of social causation and the problem with physicalistic reductionism, the biomedical model is untenable as a regulative ideal for medicine. And so, a broad biopsychosocial approach is required to accommodate the diverse range of factors involved in disease causation and to inform interventions on these factors across multiple domains.

The criticism that the biopsychosocial model is too vague to be explanatorily valuable is taken by Bolton and Gillett to suggest that the traditional version of the model lacks an appropriate account of biopsychosocial causation. Accordingly, they present a metaphysical account of biopsychosocial causation that suggests that normative processes in the biological, psychological, and social domains regulate one another through information transfer. Herein, I have raised some problems with their account and have argued that the issue of biopsychosocial causation does not have to be so metaphysically taxing, as the causal

relations between factors in the different domains can be accommodated by the more metaphysically neutral interventionist theory of causation. Furthermore, I have argued that the purported vagueness of the biopsychosocial model is not due to the issue of biopsychosocial causation, but is due to the issue of causal selection. Nonetheless, this can easily be overcome being more explicit about our explanatory interests in different contexts and more specific about the questions we ask. When this pluralistic approach to explanation is applied, the eclecticism of the biomedical model is shown not to be its weakness, but its principal strength.

Acknowledgments

I would like to thank Awais Aftab for his very generous advice on this paper and the two anonymous reviewers for taking the care to offer constructive comments. I am grateful to the Leverhulme Trust for supporting this research through an Early Career Fellowship (grant reference ECF-2017-298).

REFERENCES

Bolton, Derek, and Dinesh Bhugra. 2020. 'Changes in Society and Young People's Mental Health'. *International Review of Psychiatry*, DOI: 10.1080/09540261.2020.1753968.

Bolton, Derek, and Grant Gillett. 2019. *The Biopsychosocial Model of Health and Disease*. Cham: Palgrave Macmillan.

Broadbent, Alex. 2009. 'Causation and Models of Disease in Epidemiology'. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 40: 302–311.

Campbell, John. 2016. 'Validity and the Causal Structure of a Disorder'. In *Philosophical Issues in Psychiatry IV: Psychiatric Nosology*, edited by K. S. Kendler and J. Parnas, 257–273. Oxford: Oxford University Press.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford: Clarendon Press.

Charles, Katia, Guyomard, René, Hoyheim, Björn, Ombredane, Dominique, and Jean Luc Baglinière. 2005. 'Lack of Genetic Differentiation Between Anadromous and Resident Sympatric Brown Trout (*Salmo trutta*) in a Normandy Population'. *Aquatic Living Resources*, 18: 65–69.

Cummins, Robert. 1975. 'Functional Analysis'. *Journal of Philosophy*, 72: 741–765.

Dawkins, Richard. 1995. *River Out of Eden: A Darwinian View of Life*. New York: Basic Books.

De Vreese, Leen, Weber, Erik, and Jeroen Van Bouwel. 2010. 'Explanatory Pluralism in the Medical Sciences: Theory and Practice'. *Theoretical Medicine and Bioethics*, 31: 371–390.

Ghaemi, Nassir. 2010. *The Rise and Fall of the Biopsychosocial Model: Reconciling Art and Science in Psychiatry*. Baltimore, MD: Johns Hopkins University Press.

Griffiths, Paul E., and R. D. Gray. 1994. Developmental Systems and Evolutionary Explanation. *Journal of Philosophy*, 91: 277–304.

Hardcastle, Valerie. G. 2002. 'On the Normativity of Functions'. In *Functions: New Essays in the Philosophy of Psychology and Biology*, edited by A. Ariew, R. Cummins, and M. Perlman, 144–156. New York: Oxford University Press.

Kendler, Kenneth S. 2005. 'Toward a Philosophical Structure for Psychiatry'. *American Journal of Psychiatry*, 162: 433–440.

Lakatos, Imre. 1974. 'Science and Pseudoscience'. *Conceptus*, 8: 5–9.

Lipton, Peter. 2004. *Inference to the Best Explanation*, 2nd edition. London: Routledge.

Longino, Helen. 2013. *Studying Human Behavior: How Scientists Investigate Aggression and Sexuality*. Chicago: University of Chicago Press.

Marmot, Michael. 2005. 'Remediable or Preventable Social Factors in the Aetiology and Prognosis of Medical Disorders'. In *Biopsychosocial Medicine: An Integrated Approach to Understanding Illness*, edited by P. D. White, 39–58. New York: Oxford University Press.

Millikan, Ruth G. 1984. *Language, Thought and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

Mitchell, Sandra D. 2009. *Unsimple Truths: Science, Complexity, and Policy*. Chicago: University of Chicago Press.

Newen, Albert, De Bruin, Leon, and Shaun Gallagher. 2018. 4E Cognition: Historical Roots, Key Concepts, and Central Issues. In *The Oxford Handbook of 4E Cognition*, eds. A. Newen, L. De Bruin, and S. Gallagher, 3–18. Oxford: Oxford University Press.

Oyama, Susan. 2000. *The Ontogeny of Information: Developmental Systems and Evolution*. Durham, NC: Duke University Press.

Plomin, Robert. 2018. *Blueprint: How DNA Makes Us Who We Are*. Cambridge, MA: MIT Press.

Ratcliffe, Mathew. 2000. 'The Function of Function'. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 31: 113–133.

Ross, Lauren N. 2018. 'Causal Selection and the Pathway Concept'. *Philosophy of Science*, 85: 551–572.

Schrödinger, E. 1944. *What is Life?* Cambridge: Cambridge University Press.

Stuckler, David., and Sanjay Basu. 2013. *The Body Economic: Why Austerity Kills*. London: Allen Lane.

Woodward, John. 2004. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Zwickl, Sav, Wong, Aalex Fang, Dowers, Eden, Leemaqz, Shalem Yiner-Lee, Bretherton, Ingrid, Cook, Teddy, Zajac, Jeffrey D., Yip, Paul S. F., and Ada S. Cheung. 2021. 'Factors Associated with Suicide Attempts Among Australian Transgender Adults'. *BMC Psychiatry*, 21: 81.

EuJAP | Vol. 17 | No. 2 | 2021
UDC: 122:159.9]001
https://doi.org/10.31820/ejap.17.2.7

THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE: RESPONSES TO THE 4 COMMENTARIES

Derek Bolton[1]

[1]King's College London

Discussion – Received: 01/10/2021 Accepted: 21/10/2021

ABSTRACT

I respond to the 4 commentaries by Awais Aftab & Kristopher Nielsen (A&N), Hane Htut Maung (HHM), Diane O'Leary (DO'L) and Kathryn Tabb (KT) under 3 main headings: "What is the BPSM really?" & Why update it?; "Is our approach foundationally compromised?", and finally, "Antagonists or fellow travellers?".

Keywords: *Biopsychosocial model; causation; George Engel; information*

Preamble

First and foremost, I would like to thank the commentators—Awais Aftab & Kristopher Nielsen (A&N this issue), Hane Htut Maung (HHM this issue), Diane O'Leary (DO'L this issue) and Kathryn Tabb (KT this issue)—for the generous giving of their time to critical commentary of Derek Bolton & Grant Gillett's proposed update of the Engel's (1997) Biopsychosocial Model (B&G). I should say that while the book was co-written, this Reply is written by DB only, so the text varies between plural 'we' for the B&G book, and singular 'I' for the Reply. Our proposed update of the BPSM is in the spirit of trying to get things as straight as we can about the conceptual foundations of health, disease, and healthcare. I thank the commentators for their generous comments about the book and for their critiques on how things could be improved. There are some common and some distinctive themes in the critiques, and I will respond to them under 3 main headings: "*What Is the BPSM Really?*" & *Why Update It?*; "*Is Our Approach Foundationally Compromised?*", and finally, "*Antagonists or Fellow Travellers?*". I have aimed to include supplementary material (additional to what is in B&G) where relevant.

Correspondence: Derek.Bolton@kcl.ac.uk

1. What Was or is the BPSM Really? And Why Update It?

1.1. Was Engel Interested in Causes?

A&N highlight that biopsychosocial causation, while the main problem in B&G, was not Engel's main problem, indeed they suggest that it may not have been one of his problems at all (p. 7). At one level, this is about terminology; "causation" is semantically linked to many other expressions in the health sciences and therapeutics such as "factors" and "influences". So for example, Engel's (1977) list of what the biomedical model fails to take into account includes, quoted by A&N (p. 8-9): "for some conditions such as schizophrenia and diabetes, the effect of conditions of living on onset, presentation and course"—and we take this to refer to causal risks for onset and risk/protective factors (causally) affecting course, putting the issues squarely in the areas of epidemiology and clinical therapeutics. Another connected example, A&N propose that:

> The matters that preoccupy Engel are more to do with psychosocial influences in the form of illness interpretation and presentation, sick role, seeking or rejection of care, the doctor-patient therapeutic relationship, and role of personality factors and family relationships in recovery from illness, etc. (Aftab and Nielsen this issue, 9)

But presumably "*influences*" = something like "*make a difference to*" = "*has a causal role in*".

A&N present a convincing case that one of Engel's main and general concerns was to bring many aspects of the psychological, social dimensions of illness including the doctor-patient relationship within the realm of medical and scientific inquiry. I agree with this, but suggest that this aspiration relies on the working assumption that these dimensions are causally relevant to health outcomes of interest. This is because science, so far as I understand it, is basically concerned with causes, and this is especially so for the applied sciences that aim to make a difference. To put it briefly, healthcare will take an interest in e.g. subjective accounts of illness if it makes a difference to something relevant, e.g. to agreement on whether there is a need to treat, and how; or will take interest in social context of living if it makes (or might make) a difference to e.g. falls at home and emergency admissions; or an interest in the quality of doctor-patient communication if it affects continuing trust, attendance and acceptability of treatment; and so on. As I read Engel, much of what he says on this issue was with the intention of rejecting the dichotomy between medicine as science and medicine as 'art' (Bolton 2020).

However, this project relies on psychosocial/interpersonal factors making a difference to relevant health outcomes. In other words, this strand of Engel's BPSM is the proposal that the causal processes (factors or influences) involved in disease and healthcare are not limited to the biological, but involve the whole person in their social/interpersonal context, and, as such, they are amenable to scientific enquiry.

1.2. Was the BPSM Ever a Model?

A&N reiterate the criticisms of Nassir Ghaemi and others to the effect that the BPSM is not a model and is of no clinical or scientific value (p. 10-11). I don't want to insist that it is a model. It is probably no more of a model than the model with which Engel contrasted it, the biomedical model (BMM). Both expressions, and probably any others that summarise complex foundational issues in a word or two (such as also 'biological psychiatry', or 'phenomenology') lend themselves to various kinds of uses ranging between slogan-like and substantially theorised, with being a shorthand for a theory somewhere in between. A theorised version of the BMM would include core concepts and principles of the biomedical sciences, along with basic research and therapeutic paradigms. A theorised version of the BPSM would be the same for the biopsychosocial sciences, and this is what we attempted in B&G. We defined some core ontological and causal features of the three relevant domains and their interactions (contrast the BMM that has only one relevant domain), illustrated by some new paradigmatic biopsychosocial health-related pathways, such as those involving chronic stress and pain perception. We emphasised the theory of causal interactions between the three domains, because they are traditionally so problematic, as well as because causal explanation is central to science and its ontology.

A&N repeat Nassir Ghaemi's charge that the BPSM helped everybody to win, linked to the fact that it had no substantial scientific content (p. 10). I suspect there may be a difference here in the way that the BPSM has played out in the US and the UK. While in the US there may have been a tendency to use the BPSM as a way of being inclusive and open-minded about causes and cures, the more usual perspective in the UK seems to have been that the BPSM is more a matter of empirical data from particular studies, for example in social epidemiology and studies of stress (see e.g. White 2005). Certainly UK colleagues of mine showed some surprise at Nassir Ghaemi's interpretation of the BPSM and one UK reviewer, Julian Leff, did implicate UK/US differences (Leff 2010). This issue is probably linked to the history of "pluralism" on which more below in section 3.3.

1.3. Something's Wrong Somewhere However

Insofar as the BPSM was or has been used as a half-baked attempt at a model that served mainly to reduce uncertainty and make everybody happy, then by all means it doesn't warrant updating, just exposing and moving on. This view, however, does not sit well with the popular proposal that, nevertheless, it serves a valuable educational function, endorsed (though with apparent ambivalence) by A&N (pp. 11-13).

It seemed to us when we embarked on B&G that it was no good at all having these three propositions all being endorsed together:

(1) BPSM is the most popular model (often observed, including by HHM is his opening sentence "The [BPSM] (…) is perhaps the most widely accepted model of health and disease in contemporary medicine.")

(2) However, it is philosophically, scientifically and clinically useless—not a model at all

(3) However, it's useful in education

The combination of these three positions in the literature seemed to demand some work; doing nothing with the conjunct (1) & (2) & (3), as we saw it, was not an option.

If (2) is correct we need to abandon (1) & (3) ASAP; or we accept and retain (1) & (3), and refute or remedy (2)—and it was in this spirit of this second option that we undertook to update the BPSM.

1.4. Engel's Vision and the Value of the BPSM

At the beginning of her paper, KT uses a metaphor of psychiatry being buffeted about by centrifugal and centripetal forces, adapted from Scott Lilienfeld's paper (2014) on the DSM-5, and recognizes the potential value of the BPSM as providing a unifying, 'centripetal' force (pp. 7). KT goes on to discuss centrifugal forces in psychiatry including specialisms, by condition, by profession, by tradition and orientation. Importantly, there is sometimes conflict between specialisms, potentially leading to confusion for end users. The problem gets bigger when splitting occurs, when one side doesn't envisage the other, when there is no perceived whole, whether this be a person, healthcare, or health science. Centripetal forces, by contrast, see a conceptual unity, replacing splitting by something more holistic, and KT sees Engel's (1977) BPSM as, perhaps, the most notable

centripetal project (*loc. cit.*). I agree with that, and would add that its biggest message in this regard is not so much centripetalism within psychiatry (though this is probably an implication), but centripetalism across healthcare as a whole, positing a unity and common involvement of somatic and psychological processes.

Linked to its centripetal force, KT correctly observes that Engel's BPSM project drew on the systems theory in vogue at the time (p. 10). I suggest, however, that this was not just a sign of a temporary fashion, but was more a foretaste, a vision of what was coming: the increasing use of systems theoretic concepts and principles within and across many fields. The systems theory approach is closely linked to the acceleration of inter-disciplinary research and problem-solving programmes over recent decades, providing some general and integrating concepts and principles. In Margaret Boden's typology of interdisciplinarity, the highest levels are 'generalising' and 'integrated', involving a unified single theoretical perspective and integration around shared themes and questions (Boden 1999; see also Strijbos 2010, and Committee on Facilitating Interdisciplinary Research, Committee on Science, Engineering, and Public Policy 2004).

This is just what we were aiming at in B&G: a unified theoretical perspective and common themes (constructs and principles), relevant to health and disease, throughout the biological, psychological, and social sciences. We supposed that the BPSM could only be a truly interdisciplinary framework, able to accommodate the many kinds of factors now known to be implicated in health and disease, by having a common set of constructs and principles that operate within and between previously disparate domains. Further, we believed that, as Engel foresaw, the required set of constructs were those in systems theory, such as *function*, *design*, *ends*, *feedback*, *communication/information*, *regulation*, and *control*. Since the 1970s the systems theory approach has developed in many existing and new sciences, applied to functional structures, natural or artificial, from biology to engineering to models of social organisations, criss-crossing previous disparate domains, underpinning interdisciplinarity (see e.g. Strijbos 2010).

In fact, in the relevant recent history of ideas, there is a direct line to be traced from Schrödinger's new and original definition of life, used in B&G to characterize biology, to Engel's (1977) paper, via von Bertalanffy's General System Theory (1968). Schrödinger's work was cited by von Bertalanffy, in turn cited by Engel as a key example of the then new systems approach. Originally proposed for biology, the new systems perspectives were fast extended to cover psychological and social systems,

organised in hierarchies of complexity, from cells to societies. Engel was among those quick to recognise the relevance of these new systems perspectives to health, disease, and healthcare, along with contemporaries such as Alan Sheldon (1970), Ervin Laszlo (1972) and Howard Brody (1973). Engel used the name "biopsychosocial model" in his paper, explicitly announcing it as a new model for medicine, readily interpretable as an extension of biomedicine—and this is the name that caught on, to become now the most widely accepted model. This was a background reason for us wanting to retain the name "BPSM": the belief that its intellectual history was substantial, valid, and visionary.

By all means, along with the name came its accumulated baggage, and several colleagues and pre-publication reviewers advised that we jettison both—the name and its baggage—and propose an explicitly novel theory. However, as is well-known and noted above, the name *BPSM* is still a leading currency. We supposed that this points to the intellectual need to update it and validate the BPSM, rather than abandon it as intellectually vacuous, which is not only hard to square with its being educationally useful, but also, as suggested above, does not recognize its solid foundations.

1.5. What Moves Healthcare Mountains? Metaphysics As Continuous with Science

As noted above, KT discusses centripetal versus centrifugal forces in psychiatry, and sees the BPSM as a centripetal project, but her main concerns in her paper are the centrifugal forces that support the BMM, which she identifies as socio-economic-political (Tabb this issue, sec. 3). Given this reasonable assumption that such forces are important maintaining factors for the BMM, KT then reasonably infers that as such they are unlikely to be affected by a metaphysical argument, which she supposes to B&G to be.

In response to this I would say that the argument in B&G is not metaphysical but is meant to be scientific; actually, more accurately put, the intention is to operate in the dynamic space where metaphysics and scientific theory, and hence also data, merge. In other words, B&G buys into the idea, common in much 20th century philosophy, that philosophy (as metaphysics) is continuous with science, construed broadly as empirical knowledge. I will not spend time on this complicated issue here, but references include Quine's (1951) famous rejection of two dogmas of empiricism, and, in a different way, Lakatos' (1970) highly sophisticated philosophy of science. Importantly, metaphysics so construed is not a permanent set of truths but changes from time to time and place to place.

It undergoes major transformations, shifts in core theory (in Lakatos' 1970 terminology) or paradigm shifts (in Kuhn's 1962 terminology). This is what B&G is about, new (or relatively new) ideas in the life and human sciences that underpin the BPSM, such as Schrödinger's new characterisation of biological organisms in terms of decreasing entropy, the appearance of code in biology, AI, cognitive psychology, embodied cognition, agency, recognition of social recognition and social status *vs*. social disqualification and exclusion as processes that affect health and disease.

As this last example illustrates, interwoven with these deep theory shifts are new technologies and empirical findings, and it is these, I believe, that can move healthcare mountains—over time.

For example, I once heard the opinion that Aaron Beck and colleagues' decision to trial their new CBT for depression against meds, as being truly inspired, because, when the psychotherapy was found to outperform the pharmacotherapy (Rush et. al. 1977), it made the medical community sit up and pay attention. The data scored a reasonably direct hit on the biomedical model that envisaged biological causation only. The rest—the massively increased use of psychological therapies in healthcare systems––is recent history.

Empirical work in epidemiology has also been critical in showing the need for a broader biopsychosocial model. The new social epidemiology has shown that various forms of social exclusion, not only from biological necessities but also exclusion from psychological and psychosocial necessities, such as recognition, security, and civil rights—is bad for your health.

Here are some other, emerging candidates of research programmes closer to core biomedicine than the examples above, in cardiology and surgery. In cardiology, studies suggest that about three quarters of patients referred to rapid access cardiology clinics have non-cardiac chest pain or other symptoms, while, or but, commonly there is no management protocol for these patients and they are discharged, often to seek assessment or treatment again later (Tenkorang et al. 2006; Sekhri et al. 2007; Debney and Fox 2011; Chambers et al. 2014; Lenderink and Balkestein 2019). In surgery, there is increasing evidence that for some presentations dominated by pain, surgical procedures do not outperform placebo (Wartolowska et al. 2014; Jonas et al. 2015; Louw et al. 2017). These emerging findings appear in the context of new models of pain and subsequent new treatments. In brief, the perception and severity of pain, while typically localized in a specific part of the body, is now understood to be only partly,

and sometimes not at all, associated with local damage, but also involves higher cortical pathways processing information about the meaning and consequences of the pain for the person's life, potentially modifiable by psychosocial interventions such as psychological therapy and neuroscience education programmes (Quartana et al. 2009; Edwards et al. 2016; Andias et al. 2018). Bearing in mind that pain and associated distress and impairment of functioning are major drivers of service use, these emerging findings are of potential massive interest to healthcare provision and health economics.

To sum up, if the question is posed: what brings about major shifts in practices and great institutions such as healthcare?—then the answer is going to be complicated. Same goes for a closely related question: what kinds of factors are barriers to change? KT notes that major factors maintaining the BMM include social, cultural, economic and professional interests, noting that Engel said as much, and then infers that metaphysical considerations are unlikely to move such things. This inference looks completely right, if 'metaphysics' is understood as an exercise in the academy, in departments of philosophy, divorced from scientific theory and data. But B&G never intended this. We see the move towards a biopsychosocial framework in the health sciences, therapeutics, and epidemiology as being fundamentally a scientific paradigm shift (or series of interconnected paradigm shifts), driven by deep theory changes in combination with new empirical data. It may be that, as indicated previously (sec. 1.2.), interpreting the BPSM as a scientific project—in the broad sense including deep theory, new technologies and empirical findings—as opposed to metaphysics, or ideology, could be an interpretation more common in the UK than in the US.

KT argues for the importance of bioethics in advocating for improvements in healthcare (Tabb this issue, sec. 4) and many of her points I would agree with. I would add, however, that commonly the choice between two courses of action is based not only on the values assigned to the possible outcomes, but also on data-sensitive beliefs about how these outcomes are best likely to be achieved. Especially, whether a biomedical approach is the best way forwards or a biopsychosocial approach, or just psychosocial, will depend partly on what outcomes are desired, but also on empirical evidence about probabilities of how best to achieve them. This applies at every level, from choice of individual treatment, to choice of population level prevention programmes (options include doing nothing), to decisions on research funding priorities.

2. Is Our Approach Foundationally Compromised?

Having outlined above the intended rationale, purpose and method of B&G, the question arises whether and how far it worked out. The commentators present several major challenges to the B&G project.

2.1. Muddle about Dualism?

DO'L proposes that the BPSM always has been contradictory because on the one hand it separates the biological and the psychological, while on the other hand it rejects dualism, fudging this by inadequate definition of dualism, in the original and in B&G (pp. 8-10). She proposes that this contradiction is already in the BMM, and it transfers to the BPSM. She notes the complexity and multiple interpretations of key terms involved in defining dualism, physicalism, and reductionism (pp. 9-10).

We supposed in B&G, staying close to Engel's text, that he charged the BMM with being dualistic and committed to physicalistic reductionism. We interpreted this as meaning, briefly, that BMM is committed to ontological dualism and causal-explanatory reductionism, i.e., to the view that body and mind are ontologically distinct, but that all causing takes place at the physical level, especially that there is no causing of bodily events by mental events. This interpretation involves no contradiction between dualism and physicalist reductionism. There would be a contradiction in affirming both dualism and physicalist ontological reduction, but we don't interpret BMM as being ontologically reductionist, only causal-explanatory reductionist. The contrast is then with the BPSM, which is not explanatory reductionist, but envisages causal interactions within and between all of its three levels or domains. By all means it would be possible then to maintain that the three levels or domains were all ontologically separate, but then good luck with trying to make sense of causal interactions between them. Rather, the coherent shift is to suppose that causal interactions between the three levels of domain is possible because they are in the same ontological space, and hence our proposal that BPSM embraces the current science of embodied and embedded mind, as well as health and disease relevant aspects of the social sciences and the environmental sciences.

2.2. Clinical Utility and the "Psychosomatic" Conditions

DO'L goes on in her commentary to discuss the clinical utility of the BPSM, especially but not only for conditions that expose the unhelpful effects of dualism on healthcare, namely the so-called "psychosomatic" conditions (pp. 15-16). She expresses approval for aligning the BPSM with

evidence-based medicine. In B&G we supposed this to be now the obvious place to look for clinical guidance; substantial evidence from clinical trials and systematic reviews is available to us, unlike to Engel when he formulated the BPSM. On the other hand, DO'L criticizes B&G for placing too much faith in clinical guidance (p. 14). However, we had no intention of suggesting that clinical decision-making can be read off from clinical guidelines alone, the evidence for which is always partial, provisional, and selective (depending on the designs of the trials that have been done), without detailed history-taking and accounting for individual features of the presentation. So far as I know this crucial caveat is integral to EBM, even if there is a risk of it getting lost in practice.

However, clinical practice and the clinical studies and trials that guide it are only as good as the nosology, and as noted above, DO'L focuses particularly on the important clinical categories linked to unhelpful dualism. While there been many nosological problems and debates within physical and psychological medicine, probably none have been as conceptually problematic as those about conditions that do not fit into either of those two kinds but fall somewhere in-between. These are the called-by-many-names 'psychosomatic' conditions, themselves comprising many kinds, and, as DO'L points out, accounting for a high proportion of health conditions (p. 14). People with these conditions, associated with varying levels of distress and impairment of functioning, can be transferred between general hospitals and neurological, psychiatric or psychological clinics, too often falling between them. One aspect of this unfortunate state of affairs is the dualism that has permeated healthcare, separating the biomedical study and treatment of conditions below the neck, roughly, with neurology, psychiatry and psychology between them sharing, more or less harmoniously, the brain and mind. At the same time, the mental well-being aspects of physical health conditions have less visibility, and the same for the somatic aspects of psychiatric conditions. The continuing and probably increasing popularity of the BPSM belongs with a move towards more holistic healthcare. An important aspect of this are the new models of pain, distress and associated impairment, implicating central, not only peripheral, involvement—noted previously in section 1.5 as potentially contributing to changing healthcare practice.

2.3. Is Biological Information Still Problematic?

HHM and A&N both emphasise that the presumed normative, semantic characterization of biological information is a problematic foundation for B&G's proposed update of the BPSM. There is a substantial philosophical literature which finds such a construct problematic in biology as opposed to psychology. As A&N (p. 18) remark, we are unlikely to settle this

problem here and now, but I will summarise some aspects of the rationale why B&G proceeded in this way, and address some of the criticisms they make.

Firstly, in B&G we purposely made *regulation* and *regulatory mechanism* the primary characterization of what we suggest is a new kind of science in biology; rather than fronting the more familiar 'information-processing'. This was partly to work around the familiar philosophical objections to biological information-talk, but it was also in the belief that biology has actually moved on since the original information-processing revolution that started in the 1950s/1960s following discovery of the genetic code, and is now more involved with regulation and regulatory mechanisms throughout biological systems. These processes and mechanisms are visible: physical-chemical processes stop/start, increase/decrease; caused by observable events that lend themselves to descriptions such as 'switches' and 'gates' that e.g. increase or decrease concentration of catalysts. *Information flow* by contrast is a more abstract construct—you can't see it—and the next step of supposing that what is 'flowing' has semantic, normative content, seems to turn this abstraction into a philosophical error (horror)—at least it does when certain philosophical assumptions about content are being made, on which more below. However, as this new biological science has developed, the concept of information is not, or does not have to be seen as, doing the conceptual heavy lifting; rather it appears rolled up in a whole family of interconnected constructs, along with coding, signalling, feedback, function, and so on. This is evident in, for example, the relatively new and rapidly expanding subfields of molecular biology, cell signalling and genetic regulatory networks. As part of these developments, the construct of information is itself changing, shifting towards *programming* and *instructions*, for e.g. building complex molecules, or for the operation of regulatory mechanisms. In these theory-shifts, it is less easy to identify information-talk as having semantic content. I mean, while it is easy to assume that information is supposed to have content 'that p', where 'p' is a proposition with a truth-value expressible in language, there is no corresponding easy assumption of true/false propositional content when 'information' has the sense of *instruction*. Instructions are not true/false, though they can be e.g. normal/abnormal, or they can lead to the wrong result, in the circumstances, and they can be issued by the wrong agent. Here the reference is to the pervasive normativity in current biological models, evident in constructs such as *dysregulation*, *error*, *mutation*, *correction*, *deception/mimicry*, etc., but which is not best interpreted in terms of true/false semantic content. As to the grounds of this biological normativity, they are fundamentally to do with staying alive or dying, at the individual and/or species level.

Let me return to the point that biological semantic information or normativity is problematic only if certain philosophical assumptions about content/normativity are being made. HHM makes the criticism (p. 12), that while concepts of informational content and normativity are valid in the psychosocial domains, they are problematic in the biological domain at the sub-personal level. But apart from being familiar in folk usage, what is the metaphysics or science behind this claim? This is probably the same question as: what is the metaphysically acceptable *literal* meaning of 'informational content' and 'normativity', such that application of these terms to biological, sub-personal processes is not *literal*, but only *metaphorical*? (A&N pp. 17-18; HHM pp. 13, 15). I suggest two, completely different justifications.

One is the Cartesian or quasi-Cartesian, that would have semantic content, or intentionality and other related concepts, essentially tied to *mind* and *consciousness*. But this, I suggest, as suggested by the name of the original author, is just yesterday's science/metaphysics; the current science/metaphysics is different.

The other justification for the rejection of biological-semantic/normative talk is very different, but actually points distantly to the relevant deep shifts in science and metaphysics. It is the neo-Wittgensteinian argument, made for example by Hacker (1987), that such semantic/normativity concepts really belong to our activities using language, to language-games, i.e. briefly, to our sending/receiving signs enabling activities such as, to use an example near the start of the *Philosophical Investigations*, fetching and carrying stones for building (Wittgenstein 1953, paras. 2, 7). However, the argument in B&G is that signalling, communication, instructions, obtaining and transporting materials for building structures, is already happening in our biology—this, we contend, is the new biological science. I realise the magnitude of the alleged theory-shift here, which is basically from some idea of meaning (and cognates) as true/false representation of reality (hopefully, in Descartes), something so mysterious that only the conscious mind could do it, to the idea of meaning as communication, command and action. But this is the shift involved in the use of semantic/normative concepts in the biological as well as the psychosocial domains.

It was proposed above that the grounds of this biological normativity are fundamentally to do with staying alive or dying, at the individual and/or species level. Putting the matter thus, however, could be interpreted as grounding biological normativity in our interests and concerns, as opposed to being in independent nature. But as against that, and of course, the emergence of life on Earth and its evolution over deep time much pre-dated

us and our concerns and scientific heuristics. The difference between life and death is in nature itself, independent of us, albeit in only part of nature––the biological part.

However, Schrödinger's theory of the biological goes deeper, seeing life as dependent on building and maintaining counter-entropic dynamic structures and functions—until such time as they break down and die. It is an essential of the part of the argument in B&G, aiming to track this deep theory in current biology/biophysics, that the regularities involved in such as genetic replication, genetic regulatory mechanisms, and cell signalling, can break down. This possibility of breakdown in regularities is an essential and distinctive feature of the new biology. The biological regularities are not immutable laws of nature, like the energy exchange and conservation laws of physics and chemistry, but could be otherwise, and can fail. This refers for example to Crick's consideration of the possibility that the genetic code is a 'frozen accident', that the original allocation of codons to amino acids was "entirely a matter of 'chance'" (Crick 1968, 369-370). The accidental, non-fixed-law-like nature of the code is what allows break-down and error, as in genetic mutation, the condition of evolution, and of death.

HHM proposes (pp. 13-14) inter-linked counter-arguments to those set out in B&G, summarised above, that would distinguish biology from physics (and chemistry) in a way that permits normativity. HHM proposes that Newton's *F=ma* can lead to distinct predictions for experimental setups that are mathematically difficult to resolve. This may be true, but what is needed for to counter the argument in B&G is that *F=ma* can actually break down—and it can't. Or, it is treated in such a way that it is not allowed to break down, as in Lakatos' definitive account of scientific methodology (Lakatos 1970). Biological system-specific, information-based 'laws' always contain *ceteris paribus* clauses, as typically for the causal laws of the 'special sciences', unlike physics which has no such clauses, as argued by Fodor (1987). A statement of the sort that such-and-such genetic sequence codes for a particular protein—unpacked in terms of it producing such a protein under normal cellular operating conditions—fails to apply, breaks down, under abnormal conditions. No *ceteris paribus* clause appealing to normative conditions qualify *F=ma*.

A connected line of thought responds to HHM's connected argument (pp. 14) that teleological language can be used to describe e.g. bodies tending to thermodynamic equilibrium. But the response here is the same as applied in the massive theory-shift from Aristotelian physics to the modern mechanics of Galileo and Newton, namely, that the new non-teleological mechanics did all the work needed to explain objects falling to the ground,

and teleological language added nothing of explanatory value. In biology by contrast, the teleological language, the language of regulatory mechanisms and associated constructs, does a variety of explanatory work that is not done by physical descriptors: especially it picks out invariances among physical realisations involved with functions, tending towards ends; it identifies error and can be used to diagnose breakdown, possible repair, etc.

A specific theme in the literature endorsed by A&N (pp. 15-16) is that Shannon information is enough for biology and is not semantic. In reply to this line of thought, I would reframe but basically repeat the arguments as above: Shannon communication involves a transmitter, a signal and a receiver; information transfer reduces uncertainty in the receiver and is prone to more or less 'error'. These inter-systemic, normative concepts are quite unlike those in the energy-related laws of physics, and are applicable to artificial designed functional systems and evolved biological systems alike.

3. Antagonists or Fellow Travellers?

As befits what we argued is a large-scale theory-shift, the BPSM has many fellow-travellers, in Engel's original, and in any update now including B&G. Some among the former are mentioned in B&G, while some of the latter are cited in the commentaries as alternatives, considered below.

3.1. The Interventionist Theory of Causation a Quick Fix?

HHM argues (pp. 19-20) that the complicated and contentious causal/regulatory explanatory model proposed in B&G is not necessary to accommodate biopsychosocial causation because this can be done simply by using the interventionist theory of causation. He notes that we endorse this theory in B&G. However, I suggest, the interventionist theory is not enough by itself.

When conducting an experiment, of some degree of stringency, or by observing a natural experiment, we measure certain variables and estimate the proportion of the variance in the outcome variable that can be explained by (or at least, is associated with) different factors, using regression. It is true that we can put any measured variables that we like into the regression as independent factors, and call them 'biological', 'psychological' or 'social'. Finding that the latter two account for significant variance in health outcomes is of course a major way in which epidemiological and

clinical trials have established the evidence base for biopsychosocial models of particular health outcomes of interest.

The experimental method, however, is well known to be theory-free. So far, we have no idea of causal mechanisms, and also so far no theory of the constructs the variables stand for. In the present case, using the experimental method only, we so far have no idea how to theorise the *biological*, *psychological* or *social*—so far we just have variable names that we are saying are of these sorts. This is particularly important in this area, because of the centuries old presumptions of materialism and the consequent problematic status of psychological and social causes. In the context of this historical prejudice, apparent observations of psychosocial as well as biological causes are wide open to the reductionist pressure that would regard them as noncausal epiphenomena, which obscure the real material causes, e.g. in the brain or genes. Either way, whether we are happy with the untheorized observations, or whether we assume everything is really biological, we have no need to theorise or investigate the causal mechanisms by which e.g. psychological therapy or social exclusion affect health.

In short, the experimental method on its own, philosophically expressed as the interventionist theory of causation, delivers only sparse theory-free empirical findings. No science is satisfied with this; it requires theory, and B&G aims to articulate it for the BPSM. As discussed in B&G, the most worked out theory of how social and psychological factors impact health invokes chronic social-psychological-biological stress, and the explanatory concepts are of the sort that we try to explicate, in terms of environmental and social resources, agency, dysregulation of metabolic processes, etc. See also below section 3.3 on pluralistic approaches that include interactions between kinds of factor.

3.2. Causal Selection

HHM argues that

> the challenge when developing a defensible version of the [BPSM] (…) is not so much providing an adequate account of biopsychosocial causation, but providing an adequate account of causal selection. (Maung this issue, 21)

He notes (*loc. cit.*) that "almost every event that is caused is the outcome of multiple causal factors (...). Nonetheless, we only consider some of these causal factors to be relevant in an explanation". The issue is how we select which factors are causally relevant. HHM goes on to critically discuss

several accounts of causal selection in the literature, and in so doing covers a wide variety of considerations that may come into play in selection, ranging from empirical determination, to distinguishing between explanatorily relevant factors and background conditions, with the addition that this distinction is dependent on contexts, values, and interests, including ethical and political considerations, especially in healthcare (see Maung this issue, 21-23).

In response to this critique, I would say that while B&G does not address the question of causal selection by that name in this way, with reference to the same literature, we do come at more or less the same issues from a different angle, and arrive at quite similar conclusions. In B&G we emphasise that empirical determination is necessary to define what causes affect an outcome, and for empirical study to occur at all, a problem of interest has to have been identified, this being, in health research, a health outcome of interest—i.e. typically, a condition of range of conditions, and within that, onset, course +/- treatment, and quality of life. Once a range of causes implicated in a particular health-relevant outcome of interest has been identified, then, given that healthcare is an applied science aiming to make a difference, at the individual or population level, the challenge is to identify a causal factor that is both of *large enough effect* and is *modifiable*. Many considerations apply in all these stages: in the first step, selection of a health outcome '*of interest*', then also in decisions about what is a large-enough, modifiable target for intervention (prevention or treatment). Considerations include e.g. individual/population burden of illness; healthcare costs; acceptability of interventions, available technology, level of resources, cost-benefit analyses, political priorities—all these of different sorts. While HHM and B&G take different approaches to this question of identifying relevant causes, I don't see that they are wide apart in direction or conclusions.

3.3. Pluralism

HHM and A&N both consider the relation of the BPSM to various types of explanatory pluralism. HHM accepts that the BPSM accommodates or is compatible with explanatory pluralism (pp. 23-24), and I think that's right. A&N by contrast view explanatory pluralism as alternative to the BPSM (p. 11). On the other hand, A&N acknowledge (pp. 11, 13) that B&G's proposal that the content of the BPSM is in the specifics, is not that different to an explanatory pluralism that is guided by data on the specifics. They make the point (p. 11-12) that databased models of specific conditions, such as diabetes or depression, cannot be derived from a general statement of the BPSM, and that is of course correct and exactly part of the argument in B&G.

A&N go on to say (p. 12) that "establishing the psychological and the social as ontologically and causally real", as proposed in B&G, "doesn't help us with the question of how to best integrate the etiological factors in the form of a coherent explanation and how this should inform multidimensional approaches to treatment". My response here is that the intention in B&G is to map out, at least some of, the key constructs and principles that can be used to construct integrated models of risks for onset, maintenance, and treatment of specific conditions.

B&G considers two main models of integration: chronic stress and pain, which between them are major drivers of ill health and service use. As noted in the previous section, we highlight that current models of chronic stress are essentially biopsychosocial, involving the psychological aspect of down-regulation of agency (raising risk of dysregulation of agency, helplessness or inability to cope), interacting with the social aspect of excessive salient task demands in relation to low access to resources, linked to 'low social status', poverty, racism and other kinds of social exclusion, and the biological responses to chronic psychosocial stress that involve dysregulation of metabolic processes, compromising the immune system, creating risk for many kinds of ill health. The intention in B&G was to sketch out the constructs and principles employed in such models of complex biopsychosocial/environmental interactions. Another example considered in B&G in some detail was that of pain, discussed above in section 2.2., highlighting that current models implicate central neuropsychological processing including appraisals of agency/impairment as well as peripheral damage, or even in the absence of detectable sufficient peripheral damage. Again, the aim was to explicate the constructs and principles of these new models that integrate biopsychosocial/environmental factors.

Overall, the intention was to go beyond any general statements to the effect that "it's all very complicated involving lots of things and requiring lots of different approaches", whether such a general statement is labelled as "the BPSM" or as "pluralism". The science has gone way beyond this and there is no need for such general statements in the clinic, or in education, at least not in courses where the learning outcomes include understanding the science or the ability to read scientific papers. We can use the general statements, but hopefully followed by advice that there are ongoing research programmes on the details.

3.4. Enactivism

A&N compare and contrast the proposal in B&G with the 3/4E models of embodied cognition, sometimes called 'enactivist' theories. They note that

we endorse the 4E approach, as does HHM (p. 11), and they note many similarities between B&G and enactivism (A&N, pp. 14-15). For me, the list of similarities is long and substantial enough to regard B&G's version of the BPSM and enactivism as fellow travellers. A&N go onto contrast them, however, in favour of enactivism, citing its advantages over B&G in two respects (p. 19):

(1) Enactivism does without the problematic concept of biological normative/semantic information

(2) Enactivism explicitly bridges the natural-normative gap, by affirming that "all life shares an embodied concern (i.e. a self-perpetuating structure) for the continuation of self" (p. 19)

On the second point (2), the intention in B&G is to affirm something like what A&N propose. Specifically, and as reiterated above in section 2.3., it proposes that the biological in nature has a normativity, grounded in the difference between life and death, adding the connected point that the regularities on which life depends are contingent and mutable, unlike laws of non-biological nature, and are liable to breakdown, eventually in dying and death, the end of the struggle to withstand increasing entropy.

This raises the question of the relation between (1) and (2). Granting that enactivism envisages normativity in all life (2), why should it want to resist accepting normativity in biological information (1)? If all life exhibits normativity—grounded in the difference between life and death—what would be the problem in accepting that this normativity, so grounded, applies to biological information? It is not clear, in other words, that the first supposed advantage of enactivism sits well together with the first.

The broader point here is that models of embodied cognition such as 4E do not necessarily reject the concept of information-processing, though they of course interpret it in the terms of the model, i.e. as tied closely to requirements for action, linked to needs and concerns. What is rejected is the old idea of information-processing as being processing of 'mental representations' (Newen et al. 2018) , i.e. as I understand it, representations of a ready-made, independent world, that has so far nothing to do with the embodied, active cognitive agent. There are many strands involved in models of embodied cognition (Newen et al. 2018), and only some take the radical and problematic step of eschewing the concept of information altogether (Carney 2020).

So far as concerns the BPSM, we supposed in B&G that accounting for the biopsychological (two of the three domains in the model) requires the model of embodied cognition, which also makes explicit its essential environmental involvement. Since the BPSM also requires linkages between the psychological and social, it is also necessary to emphasise that cognition, with action and agency, is constituted by interactions not only with the non-social environment, but also by interpersonal and other social relations.

REFERENCES

Aftab, Awais, and Kristopher Nielsen. 'From Engel to Enactivism: Contextualizing the Biopsychosocial Model'. This issue. *European Journal of Analytic Philosophy*, 17 (2): (M2)5-23. https://doi.org/10.31820/ejap.17.2.3

Andias, Rosa, Neto, Maritza, and Anabela G. Silva. 2018. 'The Effects of Pain Neuroscience Education and Exercise on Pain, Muscle Endurance, Catastrophizing and Anxiety in Adolescents with Chronic Idiopathic Neck Pain: A School-Based Pilot, Randomized and Controlled Study'. *Physiotherapy Theory and Practice* 34: 682-691.

Boden Margaret A. 1999. 'What is Interdisciplinarity?'. In *Interdisciplinarity and the Organisation of Knowledge in Europe*, edited by R. Cunningham, 13-24. Luxembourg: Office for the Official Publications of the European Communities.

Bolton, Derek, and Grant Gillett. 2019. *The Biopsychosocial Model of Health and Disease: Philosophical and Scientific Developments*. London: Springer Palgrave. Open Access available at https://www.palgrave.com/gp/book/9783030118983.

Bolton, Derek. 2020. 'The Biopsychosocial Model & the New Medical Humanism'. *Archives de Philosophie* 83: 13-40. Special Issue edited by J. Ferry-Danini, and E. Giroux. Original English version at https://www.cairn-int.info/article-E_APHI_834_0013--the-biopsychosocial-model-and-the-new.htm

Brody, Howard. 1973. 'The Systems View of Man: Implications for Medicine, Science, and Ethics'. *Perspectives in Biology and Medicine* 17: 71-92.

Carney, James. 2020. 'Thinking Avant la Lettre: A Review of 4E Cognition'. *Evolutionary Studies in Imaginative Culture* 4 (1): 77-90.

Chambers, J. B., Marks, E., Knisley, L., and M. Hunter. 2013. 'Non-cardiac Chest Pain: Time to Extend the Rapid Access Chest Pain

Clinic?'. *International Journal of Clinical Practice* 67 (4): 303-306.

Committee on Facilitating Interdisciplinary Research, Committee on Science, Engineering, and Public Policy. 2004. *Facilitating Interdisciplinary Research*. Washington: National Academy Press.

Crick, Francis. 1968. 'The Origin of the Genetic Code'. *Journal of Molecular Biology* 38: 367-379.

Edwards, Robert R., Dworkin, Robert H., Sullivan, Mark D., Turk, Dennis C., and Ajay D. Wasan. 2016. 'The Role of Psychosocial Processes in the Development and Maintenance of Chronic Pain'. *The Journal of Pain* 17 (9): T70-T92.

Engel, George L. 1977. 'The Need for a New Medical Model: A Challenge for Biomedicine'. *Science* 196: 129-136.

Engel, George L. 1978. 'The Biopsychosocial Model and the Education of Health Professionals'. *Annals of the New York Academy of Sciences* 310: 169-181.

Engel, George L. 1980. 'Causation and Causal Selection in The Biopsychosocial Model of Health and Disease'. *American Journal of Psychiatry* 137: 535-544.

Fodor, Jerry. 1987. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Ghaemi, Sanjay N. 2010. *The Rise and Fall of the Biopsychosocial Model: Reconciling Art and Science in Psychiatry*. Baltimore, MD: Johns Hopkins University Press.

Hacker, Peter. 1987. 'Languages, Minds and Brain'. In *Mindwaves. Thoughts on Intelligence, Identity and Consciousness*, edited by C. Blakemore, and S. Greenfield, 485-505. Oxford: Blackwell.

Maung, Hane Htut. This issue. 'Causation and Causal Selection in the Biopsychosocial Model of Health and Disease'. *European Journal of Analytic Philosophy* 17 (2): (M5)5-27. https://doi.org/10.31820/ejap.17.2.6

Jonas Wayne B., Crawford, Cindy, Colloca, Luana, et al. 2015. 'To What Extent are Surgery and Invasive Procedures Effective Beyond a Placebo Response? A Systematic Review with Meta-analysis of Randomised, Sham Controlled Trials'. *BMJ Open* 5: e009655.

Kuhn, Thoma S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakatos, Imre. 1970. 'Falsification and the Methodology of Scientific Research Programmes'. In *Criticism and the Growth of Knowledge*, edited by I. Lakatos, and A. Musgrave, 91-196. Cambridge: Cambridge University Press.

Laszlo, Ervin. 1972. *The Systems View of the World*. New York: Braziller.

Leff, Julian. 2010. 'Review. The Rise and Fall of the Biopsychosocial Model: Reconciling Art and Science in Psychiatry. By S. Nassir Ghaemi'. *The British Journal of Psychiatry* 197: 504-505.

Lenderink, T., and E. J. M Balkestein. 2019. 'First Time Referral Reasons, Diagnoses and 10-Year Follow-Up of Patients Seen at a Dutch Fast Lane Outpatient Cardiology Clinic'. *Netherlands Heart Journal* 27: 354-361.

Lilienfeld, Scott. O. 2014. 'DSM-5: Centripetal Scientific and Centrifugal Antiscientific Forces'. *Clinical Psychology: Science and Practice* 21: 269-279.

Louw, Adriaan, Diener, Ina, Fernández-de-las-Peñas, César, and Emilio J. Puentedura. 2017. 'Sham Surgery in Orthopedics: A Systematic Review of the Literature'. *Pain Medicine* 18 (4): 736-750.

O'Leary, Diane. This issue. 'How to be a Holist Who Rejects the Biopsychosocial Model'. *European Journal of Analytic Philosophy* 17 (2): (M4)4-20. https://doi.org/10.31820/ejap.17.2.5

Newen, Alabert, De Bruin, Leon, and Shaun Gallagher. 2018. '4E Cognition: Historical Roots, Key Concepts, and Central Issues'. In *The Oxford Handbook of 4E Cognition*, edited by A. Newen, L. De Bruin, and S. Gallagher, 3-18. Oxford: Oxford University Press.

Quartana, Phillip J., Campbell, Claudia M., and Robert R. Edwards. 2009. 'Pain Catastrophizing: A Critical Review'. *Expert Review of Neurotherapeutics* 9 (5): 745-758.

Quine, Willard Van Orman. (1951). 'Two Dogmas of Empiricism'. *The Philosophical Review* 60: 20-43.

Rush, Augustus J., Beck, Aaron T., Kovacs, Maria, and Steven D. Hollon. 1977. 'Comparative Efficacy of Cognitive Therapy and Pharmacotherapy in the Treatment of Depressed Outpatients'. *Cognitive Therapy & Research* 1: 17-38.

Sheldon, A. 1974. 'Toward a General Theory of Disease and Medical Care'. *Science, Medicine and Man* 1 (4): 237-262.

Strijbos, Sytse. 2010. 'Systems Thinking'. In *The Oxford Handbook of Interdisciplinarity*, edited by R. Frodeman, K. J. Thompson, and C. Mitcham, 453–470. Oxford: Oxford University Press.

Tabb, Kathryn. This issue. 'Centrifugal and Centripetal Thinking about the Biopsychosocial Model in Psychiatry'. *European Journal of Analytic Philosophy* 17 (2): (M3)5-28. https://doi.org/10.31820/ejap.17.2.4

Von Bertalanffy, Ludwig. 1968. *General System Theory*. New York: George Braziller.

Wartolowska, Karolina, Judge, Andrew, Hopewell, Sally, Collins, Garry S., Dean, Benjamin J. F., Rombach, Ines, (…) and Andrew J. Car.

2014. 'Use of Placebo Controls in the Evaluation of Surgery: Systematic Review'. *British Medical Journal* 348: g3253.

White, Peter. ed. 2005. *Biopsychosocial Medicine, An Integrated Approach to Understanding Illness*. Oxford: Oxford University Press

Wittgenstein, Ludwig. 1953. *Philosophical Investigations*, edited by G. E. M. Anscombe and R. Rhees. Oxford: Blackwell.

INTRODUCTION TO THE SPECIAL ISSUE ON PHILOSOPHY OF MEDICINE

GUEST EDITORS

Saana Jukola[1] and Anke Bueter[2]

[1] University of Bonn
[2] Aarhus University

ABSTRACT

This article is an introduction to the special issue on philosophy of medicine. Philosophy of medicine is a field that has flourished in the last couple of decades and has become increasingly institutionalized. The introduction begins with a brief overview of some of the most central recent developments in the field. It then describes the six articles that comprise this issue.

Keywords*: philosophy of medicine; medical ethics; medical epistemology; disease; diagnosis*

1. Introduction

In the last couple of decades, philosophy of medicine has become established as a distinct branch of philosophy. While in 2008 it was possible to pose the question "Does Philosophy of Medicine Exist?" (Marcum 2008, 3), today research in the field flourishes and has become increasingly institutionalized. There are professional associations for philosophers of medicine (e.g., the Philosophy of Medicine Roundtable) and events addressing philosophical questions that arise in the context of biomedical research and clinical practice are organized regularly. In 2020 a new journal, *Philosophy of Medicine,* was established, adding to the already existing journals such as *Theoretical Medicine and Bioethics* and *Medicine, Health Care and Philosophy*. New generations of philosophers of medicine can now acquire credentials in specialized study programmes (e.g., at King's College London) and by reading introductory textbooks of philosophy of medicine (e.g., Thompson and Upshur 2017; Stegenga 2018;

Correspondence: sjukola@uni-bonn.de; abueter@cas.au.dk

Broadbent 2019). Philosophical topics are also included in the curricula in many medical schools (e.g., Tonelli and Bluhm 2020).

Research in philosophy of medicine uses tools and theoretical approaches from different areas of philosophy. Traditionally, philosophical contributions addressing medicine focused on issues either ethical or conceptual in nature (Stegenga et al. 2016). Medical ethics has millennia of history behind it and since the second half of the last century the field has become institutionalized (Jonsen 2000). Issues such as informed consent (e.g., O'Neill 2003; Beauchamp and Childress 2006), euthanasia (e.g., Rachels 2019) and questions related to justice regarding the access to healthcare (e.g., Daniels 2001; Powers and Faden 2006) have been discussed in journals and conferences dedicated to the field. Conceptual explorations related to medical practice have, in turn, typically focused on the definitions of 'disease' and 'health' (e.g., Boorse 1977; Cooper 2002).

During the last years, contributions to medical epistemology have grown in number. Questions concerning, for instance, evidential standards used for evaluating causal claims or problems related to clinical decision-making have become more central. In particular, the development and pre-eminence of evidence-based medicine has sparked a lively debate about which methods should be used for making claims about the effectiveness of different interventions. Scholars have been especially interested in presenting arguments for and against the use of randomized controlled trials in comparison to other ways of collecting evidence (e.g., Howick 2011; Parkkinen et al. 2018). With respect to clinical practice, a prominent question has been how evidence, expertise and patient values should be integrated into decision-making (e.g., Tonelli 2006; Loughlin et al. 2017). The use of artificial intelligence in the clinical context is another emerging focus of research (e.g., Genin and Grote 2021). Moreover, the experiences of patients and the epistemic status of their testimonials have been analysed by drawing on Miranda Fricker's (2007) work on epistemic injustice. For instance, Carel and Kidd (2014) have argued that ill persons in general face testimonial and hermeneutical injustices, a problem even more prominent for patients with mental illnesses (e.g., Bueter 2019, 2021; Crichton et al. 2017; Scrutton 2017). A related addition to the conceptual debate on health and disease is the phenomenology of illness that focuses on the lived experience of patients (e.g., Carel 2011, 2016; Ratcliffe 2014). This focus on patient perspectives can, in turn, impact our thinking about the study, classification, and treatment of diseases.

Social epistemology has turned out to be a particularly fruitful tool for analysing how institutional and social factors influence research and practice in different areas of healthcare. For example, the impact of

commercial interests in pharmaceutical research has attracted ample attention (e.g. Biddle 2007; Holman 2019; Bueter and Jukola 2020). In addition, the problem of neglected diseases has inspired scholars to apply theories from political philosophy to the evaluation of the distribution of research efforts in biomedical sciences (e.g. Reiss and Kitcher 2009). Besides such economic and institutional matters, scholars have noted that the social context can also affect medical research by introducing value-laden background assumptions and concepts. For example, this relates to categories of race and gender and the question whether and how these should be treated as significant variables in health science research (e.g., Bueter 2017; Valles 2021).

Metaphysical questions studied by philosophers of medicine include, for example, the nature of the relationship between pregnant organisms and fetuses (Kingma 2019) and the question of diseases as natural kinds (Beebee and Sabbarton-Leary 2010). Ontological commitments in mainstream biomedicine have been discussed by Marcum (2008), among others.

Another notable development in philosophy of medicine is the growing interest in epidemiology. Epidemiological research has attracted philosophers' attention since Alex Broadbent's seminal book (Broadbent 2013). During the COVID-19 pandemic many philosophers have increasingly focused on, for example, the epistemic nature of theories, causal inference and data practices in epidemiology—often publishing together with scholars from other fields (e.g., Broadbent et al. 2020; Fuller 2021; Harvard et al. 2021). The interconnectedness of ethical and epistemic aspects of research (for instance to health disparities) is another area where philosophers of medicine have contributed to the study of epidemiology (e.g., Katikireddi and Valles 2015; Amoretti and Lalumera 2020).

As noted by Thaddeus Metz and Chadwin Harris (2018, 282), philosophers of medicine have typically drawn on Western medical sources while overlooking healthcare practices in other parts of the globe. However, some scholars have addressed other medical practices. In their article, Metz and Harris discuss some fruitful philosophical questions that arise from African sources. Lee (2017), in turn, addresses philosophical foundations of Chinese medicine.

2. Papers in the Special Issue

An important motivation for this special issue was the observation that many of the particularly critical philosophical questions that arise in the context of healthcare cannot be answered by drawing on one philosophical tradition alone. Traditionally there has been a gap between, for example, bioethics and medical epistemology, and contributions to these fields have been published and discussed in different fora. However, as the COVID-19 pandemic made clear, and as all of the articles in this special issue show, ethical, socio-political, epistemic and ontological issues in philosophy of medicine are often deeply interconnected. For instance, the question of what mitigation measures should be undertaken to control the pandemic cannot be answered without considering both the effectiveness of the measures in slowing the spread of the virus and their political implications. Similarly, the classification of diseases gives rise to problems that are at the same time epistemic, ethical, and political.

Ashley Graham Kennedy and **Bryan Cwik** delve into issues related to diagnostic testing in the COVID-19 pandemic. Diagnosis, as they emphasize, is an essential cornerstone of clinical medicine. As such, it deserves more attention from philosophers of medicine, as it gives rise to a host of ethical and epistemic questions. Kennedy and Cwik develop a concept of diagnostic justice as requiring an equitable distribution of the burdens and benefits of testing. Looking at COVID-19 through this lens of diagnostic justice, they differentiate three areas in which testing is undertaken: in the clinical care for individuals, as an entry criterion for trials in clinical research, and in surveillance on the population level. These areas come with different goals for testing, which need to be clearly communicated and give rise to ethical questions about the moral obligations towards test subjects in these specific contexts.

Philosophical questions raised by the COVID-19 pandemic are also addressed by the second paper in this special issue. In her article, **Daria Jadreškić** looks at adaptive clinical trials. In contrast to fixed randomized controlled trials, these allow for changes of design features during a trial, based on interim results. While this comes with an increased risk of certain biases, adaptive design trials also have advantages such as a faster proliferation of results. Unsurprisingly, they have therefore played a big role in pandemic research—from Ebola to COVID-19. Jadreškić argues that adaptive design trials do not in principle lack validity. Rather, validity has to be assessed on a case by case basis (as with fixed randomized controlled trials) and with a focus on operational conditions and implementation. In addition, she shows that adaptive trial design is not a novelty introduced by COVID-19 research, but can be placed within the larger

context of the productivity crisis in pharmaceutical research and new developments in translational medicine.

Anne-Marie Gagné-Julien's paper contributes to the burgeoning literature on pathocentric epistemic injustices. She argues that the framework of epistemic injustice can be fruitfully applied to the question of how to identify wrongful medicalization. Rather than focusing on a substantive account of medicalization, which aims to tie the legitimacy of medicalization to, e.g., the presence of harmful dysfunction, she takes her departure from Kaczmarek's pragmatic account of medicalization. She proposes to expand this account with a focus on epistemic injustices created or diminished by specific procedures instrumental in medicalization. She then applies this to the case of "Premenstrual Dysphoric Disorder", a diagnosis added to the Diagnostic and Statistical Manual of Mental Disorders (DSM) in 2013. Here, the focus on epistemic injustice shows why this is a problematic case of medicalization, as the process of the diagnosis' establishment lacked in inclusivity.

Medicalization is also at the heart of **Jacob Stegenga's** contribution, which deals with yet another gender-specific disease category, namely low female sexual desire. The respective DSM diagnosis of "Female Sexual Interest/Arousal Disorder" has stirred a lot of controversy, not least because of the recent approval of pharmaceutical treatments. Stegenga identifies two major and conflicting perspectives on low female sexual desire. The mainstream view considers it a genuine disease and often focuses on biological underpinnings of low levels of desire, as well as on pharmaceutical solutions. By contrast, the critical view focuses on the social context and cultural factors that impact sexuality and respective ideas of normality. Stegenga analyzes the main arguments for each camp––which include disagreements on empirical as well as normative issues—and proposes to focus on pragmatic considerations of the harms and benefits of medicalization.

Kathleen Murphy-Hollies applies Jerome Wakefield's concept of mental disorder as harmful dysfunction (HD) to the case of gender dysphoria. She argues that HD fails to reach its own goal of avoiding a pathologization of normal states, because it leaves the relation between its components ("harm" and "dysfunction") undertheorized. She argues that we have to take a closer look at why exactly purported dysfunctions in gender dysphoria are perceived as harmful and disvalued. Firstly, this leads her to a distinction between sex dysphoria and gender dysphoria, that correlate with different sources of dysfunction and harm. Secondly, she shows that the legitimacy of the diagnosis of gender dysphoria depends on how we conceptualize gender in a sociological sense, thereby calling for a greater

involvement of sociological theory in discussions of (gendered) medicalization issues.

Thomas Schramme approaches the underlying issues in the problem of medicalization from a more general and conceptual angle. His paper addresses the problem of how to draw a line between "functional" and "dysfunctional" in functions that allow for grades. This quantitative problem of where to draw a threshold has recently played a big role in the debate on normativist versus naturalist conceptions of disease. Schramme argues that the quantitative problem does not require us to make value-laden or arbitrary decisions, but can be based on biological facts about goal-effectivity. Thus conceived, biological dysfunction is a necessary condition for a state or process to be a disease. Yet it is not sufficient, as Schramme shows by introducing a distinction between biological and clinical dysfunction. While the identification of clinical dysfunction calls for evaluative and pragmatic considerations, the fact that it is based on empirical questions about biological functions helps to avoid over-medicalization, as Schramme argues.

REFERENCES

Amoretti, Maria Cristina, and Elisabetta Lalumera. 2020. "The Concept of Disease in the Time of COVID-19." *Theoretical Medicine and Bioethics* 41 (5–6): 203–21. https://doi.org/10.1007/s11017-021-09540-5.

Beauchamp, Tom L., and James F. Childress. 2006. *Principles of Biomedical Ethics*. 6th ed. New York: Oxford University Press.

Beebee, Helen, and Nigel Sabbarton-Leary. 2010. "Are Psychiatric Kinds Real?" *European Journal of Analytic Philosophy* 6 (1): 11–27.

Biddle, Justin. 2007. "Lessons from the Vioxx Debacle: What the Privatization of Science Can Teach Us about Social Epistemology." *Social Epistemology* 21 (1): 21–39. https://doi.org/10.1080/02691720601125472.

Boorse, Christopher. 1977. "Health as a Theoretical Concept." *Philosophy of Science* 44 (4): 542–73. https://doi.org/10.1086/288768.

Broadbent, Alex. 2013. *Philosophy of Epidemiology*. New Directions in the Philosophy of Science. Basingstoke, Hampshire ; New York: Palgrave Macmillan.

———. 2019. *Philosophy of Medicine*. New York, NY: Oxford University Press.

Broadbent, Alex, Herkulaas Combrink, and Benjamin Smart. 2020. "COVID-19 in South Africa." *Global Epidemiology* 2 (November): 100034.

https://doi.org/10.1016/j.gloepi.2020.100034.

Bueter, Anke. 2017. "Androcentrism, Feminism, and Pluralism in Medicine." Topoi 36 (3), 521-530. https://doi.org/10.1007/s11245-015-9339-y

———. "Epistemic Injustice and Psychiatric Classification." *Philosophy of Science* 86 (5): 1064–74. https://doi.org/10.1086/705443.

———. 2021. "Diagnostic Overshadowing in Psychiatric-Somatic Comorbidity: A Case for Structural Testimonial Injustice." *Erkenntnis*, 1–21. https://doi.org/10.1007/s10670-021-00396-8.

Bueter, Anke, and Saana Jukola. 2020. "Sex, Drugs, and How to Deal with Criticism: The Case of Flibanserin." In *Uncertainty in Pharmacology*, edited by Adam LaCaze and Barbara Osimani, 338:451–70. Boston Studies in the Philosophy and History of Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-29179-2_20.

Carel, Havi. 2011. "Phenomenology and Its Application in Medicine." *Theoretical Medicine and Bioethics* 32 (1): 33–46. https://doi.org/10.1007/s11017-010-9161-x.

———. 2016. *Phenomenology of Illness*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199669653.001.0001.

Carel, Havi, and Ian James Kidd. 2014. "Epistemic Injustice in Healthcare: A Philosophial Analysis." *Medicine, Health Care and Philosophy* 17 (4): 529–40. https://doi.org/10.1007/s11019-014-9560-2.

Cooper, Rachel V. 2002. "Disease." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 33 (2): 263–82. https://doi.org/10.1016/S0039-3681(02)00018-3.

Crichton, Paul, Havi Carel, and Ian James Kidd. 2017. "Epistemic Injustice in Psychiatry." *BJPsych Bulletin* 41 (2): 65–70. https://doi.org/10.1192/pb.bp.115.050682.

Daniels, Norman. 2001. "Justice, Health, and Healthcare." *American Journal of Bioethics* 1 (2): 2–16. https://doi.org/10.1162/152651601300168834.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Fuller, Jonathan. 2021. "What Are the COVID-19 Models Modeling (Philosophically Speaking)?" *History and Philosophy of the Life Sciences* 43 (2): 47. https://doi.org/10.1007/s40656-021-00407-5.

Genin, Konstantin, and Thomas Grote. 2021. "Randomized Controlled Trials in Medical AI: A Methodological Critique." *Philosophy of Medicine* 2 (1). https://doi.org/10.5195/philmed.2021.27.

Harvard, Stephanie, Eric Winsberg, John Symons, and Amin Adibi. 2021. "Value Judgments in a COVID-19 Vaccination Model: A Case Study in the Need for Public Involvement in Health-Oriented Modelling." *Social Science & Medicine* 286: 114323. https://doi.org/10.1016/j.socscimed.2021.114323.

Holman, Bennett. 2019. "Philosophers on Drugs." *Synthese* 196 (11): 4363–90. https://doi.org/10.1007/s11229-017-1642-2.

Howick, Jeremy. 2011. *The Philosophy of Evidence-Based Medicine*. Chichester, West Sussex, UK: Wiley-Blackwell, BMJ Books.

Jonsen, Albert R. 2008. *A Short History of Medical Ethics*. Oxford: New York: Oxford University Press.

Katikireddi, S. Vittal, and Sean A Valles. 2015. "Coupled Ethical–Epistemic Analysis of Public Health Research and Practice: Categorizing Variables to Improve Population Health and Equity." *American Journal of Public Health* 105 (1): e36–42. https://doi.org/10.2105/AJPH.2014.302279.

Kingma, Elselijn. 2019. "Were You a Part of Your Mother?" *Mind* 128 (511): 609–46. https://doi.org/10.1093/mind/fzy087.

Lee, Keekok. 2017. *The Philosophical Foundations of Classical Chinese Medicine: Philosophy, Methodology, Science*. Lanham: Lexington Books.

Loughlin, Michael, Robyn Bluhm, Stephen Buetow, Kirstin Borgerson, and Jonathan Fuller. 2017. "Reasoning, Evidence, and Clinical Decision-Making: The Great Debate Moves Forward." *Journal of Evaluation in Clinical Practice* 23 (5): 905–14. https://doi.org/10.1111/jep.12831.

Marcum, James A. 2008. "Reflections on Humanizing Biomedicine." *Perspectives in Biology and Medicine* 51 (3): 392–405. https://doi.org/10.1353/pbm.0.0023.

Metz, Thaddeus, and Chadwin Harris. 2018. "Advancing the Philosophy of Medicine: Towards New Topics and Sources." *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 43 (3): 281–88. https://doi.org/10.1093/jmp/jhy007.

O'Neill, Onora. 2003. "Some Limits of Informed Consent." *Journal of Medical Ethics* 29 (1): 4–7. https://doi.org/10.1136/jme.29.1.4.

Parkkinen, Veli-Pekka, Christian Wallmann, Michael Wilde, Brendan Clarke, Phyllis Illari, Michael P Kelly, Charles Norell, Federica Russo, Beth Shaw, and Jon Williamson. 2018. *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*. SpringerBriefs in Philosophy. Cham: Springer

International Publishing. https://doi.org/10.1007/978-3-319-94610-8.

Powers, Madison, and Ruth R. Faden. 2006. *Social Justice: The Moral Foundations of Public Health and Health Policy*. Issues in Biomedical Ethics. Oxford, New York: Oxford University Press.

Rachels, James A. 2019. "Active and Passive Euthanasia." In *The Social Medicine Reader, Volume I, Third Edition*, edited by Jonathan Oberlander, Mara Buchbinder, Larry R. Churchill, Sue E. Estroff, Nancy M. P. King, Barry F. Saunders, Ronald P. Strauss, and Rebecca L. Walker, 273–79. Duke University Press. https://doi.org/10.1215/9781478004356-042.

Ratcliffe, Matthew. 2015. *Experiences of Depression: A Study in Phenomenology*. First edition. International Perspectives in Philosophy and Psychiatry. Oxford: Oxford University Press.

Reiss, Julian, and Philip Kitcher. 2009. "Biomedical Research, Neglected Diseases, and Well-Ordered Science." *Theoria. Revista de Teoría, Historia y Fundamentos de La Ciencia* 24 (3): 263–82.

Scrutton, Anastasia Philippa. 2017. "Epistemic Injustice and Mental Illness." In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, Josè Medina, and Gaile Pohlhaus. London: Routledge.

Stegenga, Jacob. 2018. *Care and Cure: An Introduction to Philosophy of Medicine*. Chicago: The University of Chicago Press.

Stegenga, Jacob, Ashley Kennedy, Serife Tekin, Saana Jukola, and Robyn Bluhm. 2016. "New Directions in Philosophy of Medicine." In *Bloomsbury Companion to Contemporary Philosophy of Medicine*, edited by James Marcum, 343–67. London: Bloomsbury Academic.

Thompson, R. Paul, and Ross Upshur. 2018. *Philosophy of Medicine: An Introduction*. London, New York: Routledge, Taylor & Francis Group.

Tonelli, Mark R. 2006. "Integrating Evidence into Clinical Practice: An Alternative to Evidence-Based Approaches: Integrating Evidence into Clinical Practice." *Journal of Evaluation in Clinical Practice* 12 (3): 248–56. https://doi.org/10.1111/j.1365-2753.2004.00551.x.

Tonelli, Mark R., and Robyn Bluhm. 2020. "Teaching Medical Epistemology within an Evidence-Based Medicine Curriculum." *Teaching and Learning in Medicine* 33 (1): 98–105. https://doi.org/10.1080/10401334.2020.1835666.

Valles, Sean A. 2021. "Why Race and Ethnicity Are Not like Other Risk Factors: Applying Structural Competency and Epistemic Humility in the Covid-19 Pandemic." *Philosophy of Medicine* 2 (1). https://doi.org/10.5195/philmed.2021.52.

European Journal of Analytic Philosophy EuJAP | Vol. 17 | No. 2 | 2021
https://doi.org/10.31820/ejap.17.3.1

DIAGNOSTIC JUSTICE: TESTING FOR COVID-19

Ashley Graham Kennedy[1] and Bryan Cwik[2]

[1] Florida Atlantic University
[2] Portland State University

Original scientific article – Received: 07/04/2021 Accepted: 13/08/2021

ABSTRACT

Diagnostic testing can be used for many purposes, including testing to facilitate the clinical care of individual patients, testing as an inclusion criterion for clinical trial participation, and both passive and active surveillance testing of the general population in order to facilitate public health outcomes, such as the containment or mitigation of an infectious disease. As such, diagnostic testing presents us with ethical questions that are, in part, already addressed in the literature on clinical care as well as clinical research (such as the rights of patients to refuse testing or treatment in the clinical setting or the rights of participants in randomized controlled trials to withdraw from the trial at any time). However, diagnostic testing, for the purpose of disease surveillance also raises ethical issues that we do not encounter in these settings, and thus have not been much discussed. In this paper we will be concerned with the similarities and differences between the ethical considerations in these three domains: clinical care, clinical research, and public health, as they relate to diagnostic testing specifically. Via an examination of the COVID-19 case we will show how an appeal to the concept of diagnostic justice helps us to make sense of the (at times competing) ethical considerations in these three domains.

Keywords: *diagnostic justice; philosophy of medicine; political philosophy; applied ethics*

Correspondence: kennedya@fau.edu

1. Introduction

The ongoing SARS-CoV-2/COVID-19 pandemic, now (August 2021) over 18 months old, has proved to be the greatest public health challenge and most significant global health event since the 1919 H1N1 influenza pandemic. This is so not just because of the scale, devastation, and human toll of the pandemic, but also because of some of the unique features of the disease itself. As has been well-documented, COVID-19 disproportionately causes severe illness among older adults, especially older males with certain underlying health conditions. The disease has entered the world at a unique time in human history, when large portions of the population are older and have age-related chronic conditions such as renal disease, diabetes, and hypertension, meaning that many more living individuals are susceptible to severe outcomes from this virus in a way that wouldn't have been the case a generation ago (Onder et al. 2020; Begley 2020). It has also exposed an existing and pernicious set of underlying, unjust inequalities, resulting in a distribution of mortality and morbidity that disproportionately impacts communities of color and low-income workers in developed countries (Hooper et al. 2020), as well as long-standing, pernicious inequalities in health care provision and access to medicines that exist between developed and developing countries.

One of the major challenges of the pandemic has been diagnostic testing for SARS-CoV-2 infection. Because of the danger of asymptomatic and pre-symptomatic transmission, testing is required in order to bring transmission of the disease under control, as it is the primary way in which to identify asymptomatic or pre-symptomatic cases and thus to control transmission via isolation of these individuals (Furukawa et al. 2020). Countries that have done well with testing (such as South Korea and Singapore) have fared better than other countries where testing has been more limited, such as the United States (Cheng et al. 2020). But testing in the context of this pandemic is, as in medicine and health care practice more generally, done for different purposes, and sorting through the rationale for COVID-19 testing, its different uses, and its relevance in different settings is a major conceptual and normative issue raised by the pandemic and the public health response to it.

Even aside from the COVID-19 considerations we will examine in detail here, it is not an overstatement to say that that the process of diagnosis—of which testing for infectious disease is an element—is the cornerstone of modern clinical medicine. This is because before the treatment or prognostic evaluation of any patient can begin, there must be at least a working diagnosis—some idea of what is causing the problem that brought the patient into the clinic in the first place. If a clinician does not begin the

clinical encounter by working to obtain an accurate, or at least close to accurate, diagnosis, then subsequent treatments prescribed for the patient are likely to be ineffective, and prognoses to be inaccurate. This means that clinicians must be concerned with the questions of when, how, and why to test their patients in order to best facilitate their individual health outcomes.

But diagnostic testing also has purposes beyond that of facilitating the clinical care of individual patients: it can also be used as an inclusion criterion for clinical trials, or in certain cases to surveil for, contain, and/or mitigate disease. In these cases, the goals of the testing are different from those of clinical care, and so are the ethical issues that arise when testing is conducted in these other domains. All of these different purposes for testing are present in the context of the COVID-19 pandemic, but they are not always carefully separated, and the running together of testing for clinical care and surveillance, in particular, has raised some important ethical and philosophical difficulties.

In this paper we will consider some of these difficulties via an exploration of the concept of *diagnostic justice* (Kennedy 2021) in the context of the COVID-19 pandemic, by examining the overlapping categories and the philosophical issues that arise out of diagnostic testing for clinical trial inclusion, public health surveillance, and testing to facilitate the clinical care of individual patients. In particular, we will focus on two areas of difficulty that require closer scrutiny: the possibility that individuals could confuse the goals of testing for public health surveillance with testing for clinical care, and the way that testing data is used to inform public health decision-making. We will argue that both of these areas raise issues of diagnostic justice regarding how testing is conducted and how testing data is utilized in managing the pandemic.[1] Our aim here is to point out two areas of difficulty that require further investigation and fine tuning of testing policy in the future. The COVID-19 pandemic is still, as of the writing of this paper, very much underway, and there remains much to be learned about the global response to it. This paper is thus written in the spirit of raising some questions that deserve reflection and analysis as the entire world endeavors to understand what has happened (and is happening) during this period, and to prepare for future global health emergencies.

[1] We refrain here from offering any judgment on whether testing policy for COVID-19 has failed to meet demands of diagnostic justice. The situation is still emergent, and we believe a sober judgment will need to be made retrospectively, once the pandemic is under control and there is more evidence available. We thank an audience at Georgetown University, for pushing us to clarify our aims here.

In the following section we will survey the different forms of testing for COVID-19 and then in section 3 we will outline some of the ethical issues that arise when these testing methods are employed. In section 4 we will discuss the idea of diagnostic justice and argue that issues of justice are generated by the uses of diagnostic testing in different settings. In section 5 we will raise two ethical difficulties regarding diagnostic justice for COVID-19 testing. We will then draw out some implications of this discussion for diagnostic justice, testing, and global public health policy in section 6, before a brief conclusion in section 7.

2. COVID-19 Testing Methods

Types of tests

There are three main types of tests currently in use for the diagnosis/detection of COVID-19 infection. Two of them (PCR testing and antigen testing) are used to detect active infection, while the third (antibody testing) is used to detect past infection with the SARS-CoV-2 virus. The PCR test for COVID-19 infection is considered to be highly accurate, but at this time no data on the exact sensitivity or specificity of the test is available, because there is no gold standard to compare it to. However, estimates based on similar PCR tests for other diseases put the specificity of the COVID-19 test very high (close to 100 percent, barring lab or technician error), but sensitivity only at around 70 percent, due to the relative frequency of inadequate sampling as well as the disease's variable incubation period (estimated as 2-14 days). Antigen testing, on the other hand, has the benefit of delivering results quickly (usually in about 15-20 minutes), which can be useful in point-of-care treatment for patients, but it is less sensitive than PCR testing and thus delivers more false negative results.

Antibody testing, in contrast to PCR and antigen testing, is used to confirm a past infection with the SARS-CoV-2 virus. Because measuring antibody levels in a large segment of the population can help to determine how much of the population is or was infected with the virus, which in turn allows for an estimation of the level of herd immunity present in that population, antibody testing can be very useful for public health surveillance. Of course, measuring antibody levels in a population in order to estimate herd immunity is useful only if naturally derived antibodies do indeed provide immunity to the disease. Given preliminary data, this does seem to be a reasonable assumption (Spellberg et. al. 2021) in the case of COVID-19.

Test Uses

In the clinical setting, COVID-19 testing is conducted on individuals for the purpose of diagnosing those patients who are either symptomatic, or who have had recent exposure to the virus, in order to facilitate their individual case management. In the context of a research trial, on the other hand, potential participants are tested as an inclusion criterion for the trial, in order to make sure that symptoms are due to COVID-19, rather than other respiratory infections or disorders. In the public health domain, there are at least three reasons why a COVID-19 test might be conducted: for screening, for passive surveillance, or for active surveillance. According to the CDC,

> The primary purpose of screening is to identify early signs and symptoms of a disease or health problem to implement early treatment or program intervention to reduce the likelihood of the emergence of disease or health problem and/or mortality from the disease in an individual. (Oleske 2009, 131)

So far, COVID-19 tests have not been used for this purpose, although it is possible that in the future, especially if early treatment or prevention measures become available, that they might be. COVID-19 tests can also be used for the purpose of passive surveillance, which "is intended to monitor community- or population-level outbreak of disease, or to characterize the incidence and prevalence of disease" (Center for Disease Control and Prevention 2020). Surveillance testing is performed on de-identified specimens, usually via antibody titer on samples obtained from clinics or hospitals, and thus the results are not linked to individual patients or participants. Because of this, surveillance testing cannot be used for individual patient care, however it is often used as decision-input for population level health interventions (Oleske 2009). The sort of testing for COVID-19 that is most often conducted in the public health domain is for the purpose of active surveillance. Confusingly, sometimes the literature (and the CDC) refers to this also as "screening". However, the purpose of this kind of testing is different than screening, because the goal is not to treat or prevent disease in individuals, but rather to

> identify infected persons who are asymptomatic and without known or suspected exposure to SARS-CoV-2. [It] is performed to identify persons who may be contagious so that measures can be taken to prevent further transmission. (Oleske 2009, 139)

In practice, however, this theoretically strict separation of goals often becomes blurred, and both participants in trials and the researchers that conduct them are forced to navigate potentially complicated situations. As an example, consider the role of testing in AIDS vaccine trials. Testing during AIDS vaccine field trials is essential in order to collect data on the efficacy of vaccine candidates. There is, quite simply, no way to know whether a vaccine is working or not without the testing of the subjects in the trial. Further, because of the manner of presentation and progressive nature of the disease, testing for HIV infection is necessary for the diagnosis of AIDS. What this means in practice is that while subjects can of course refuse to participate in the trial altogether, or to withdraw from the trial at any time, they cannot refuse testing and at the same time remain in the trial; if they are not able to consent to testing, then they cannot participate. However, during AIDS vaccine trials, testing also often ends up serving a de facto clinical function. Because these trials are mostly staged in developing countries with high baseline transmission rates, or in populations with a high risk of HIV infection, there is a significant chance that, even despite counseling, provision of different services, and of course some individuals getting the vaccine candidate itself, individuals in (but not only in) control groups will become HIV positive. There has been a longstanding debate about the obligations researchers have to subjects in these trials who become HIV positive during the course of the research (Berkley 2003). It is now generally accepted that researchers have *some* obligations to provide some form of care and support for HIV positive research subjects enrolled in clinical trials for HIV/AIDS therapeutics, such as the provision of antiretroviral medication and financial support for health infrastructure in communities from which participants are drawn (Richardson 2007). This means that in the course of conducting diagnostic testing for HIV infection for research purposes, data from this testing also has a clinical function, in that it identifies individuals that are (potentially) owed some form of care as part of the duty researchers owe to participants. So, while superficially similar to the ethical issues involved with diagnostic testing in clinical care, testing as part of clinical research raises different concerns.

Public Health

Diagnostic testing for public health reasons is subject to a seemingly similar issue as is testing that is used in the context of clinical research, in that its primary goal is not (necessarily) to benefit the individuals submitting to the testing, but rather to protect the public health as a whole. But, as in the case of clinical research, there is, in practice, often a blurring of these goals. For example, submitting to testing to provide pieces of aggregate data for public health purposes can also have an important

clinical benefit for individuals, as it allows them to also provide information to their providers that can help to facilitate their own care. However, this blurring of clinical medicine vs. public health raises some difficulties for the ethics of COVID-19 testing, which we will discuss in section 5 below.

When it comes to the question of whether individuals can refuse testing for public health purposes, the situation is far murkier than it is with clinical research. With passive surveillance, individuals can refuse testing without compromising the public health goals of collection of data, as long as there is a sufficient sample who will submit to testing (or some form of proxy data that can be gathered instead). But with active surveillance, the situation is different. This sort of testing, for example, is often required for things like crossing borders where mandatory quarantine orders or travel restrictions are in effect. Refusing to submit to testing in this kind of context can be grounds for the barring of entry or even for forcing individuals into mandatory quarantine. Active surveillance requires a high volume of testing; during the COVID-19 pandemic, different countries have taken different tacks when it comes to mandating testing during active surveillance. Though compelling testing (as in China) raises some serious ethical questions, leaving testing voluntary (as has been the case in the United States) raises its own difficulties (which we will also discuss in section 5 below).

There is an enduring question here about whether testing for public health surveillance can be compelled. On the one hand, there is a clear public health rationale based on prevention of harms to others for making testing mandatory, at least in certain circumstances.

On the other hand, as we will argue in the next two sections, the way testing data is used is not morally inert. Compelling individuals to submit to testing, and then using data in ways that either results in an inequitable distribution of the burdens of mitigation or neglects obligations of care to individuals would raise serious concerns. Whether compelling testing is justifiable, then, depends on a number of factors. Some of these factors are unique to the situation of testing for disease surveillance in public health, and some are shared with other domains in which diagnostic testing is employed (as we've noted, with testing for clinical research, where compelling testing as a condition of participation also raises questions about ancillary duties of care).[2] So, the ethics of diagnostic testing for an

[2] We offer here no opinion on whether testing for COVID-19 in situations where it was left voluntary (such as in community testing in the United States) should have been mandatory. No general opinion

infectious disease such as COVID-19, while it raises some common questions in all scenarios (such as questions about a right to refuse a test as well as about balancing different goals of testing), is sensitive to differences in context between clinical care, clinical research, and public health settings. Understanding these differences is crucial to understanding the concerns of diagnostic justice raised by testing for public health purposes.

3. Diagnostic Justice

In biomedical ethics much has been written about the idea of justice as fairness, particularly as it relates to the allocation of treatments to patients, especially when these treatments are scarce resources in the community (Beauchamp and Childress 2020; Emanuel, et. al. 2020; Truog et. al 2020). However, at least to our knowledge, this concept has not been discussed in regard to diagnostic testing. It is our view, however, that in the case of diagnostic testing, as with health care generally, there are multiple, and sometimes competing, moral considerations that come into play when making decisions about allocating testing resources, using data, and compelling (or not compelling) individuals to submit to testing. In some instances, there are not enough diagnostic tests to go around (as was the case in the early days of the COVID-19 pandemic in the United States), while in other cases, even when there is an adequate supply of tests, the act of testing itself can have differential impacts on the individuals being tested (this is further discussed in section 5, below) and thus there arise distributive considerations in how testing should be used and what resources should made available to those who submit to testing. In our view, what this means is that diagnostic testing is subject to demands of *diagnostic justice* (Kennedy 2021). That is, diagnostic justice requires both that the burdens and benefits of testing be distributed equitably and that diagnostic resources be allocated fairly. Thus, diagnostic justice, like other forms of justice,

is possible, as the rationale for compelling testing is sensitive to highly local factors—any justification for compelling testing will depend at least to some degree on how much harm results from a voluntary testing regime, and this will always be something that must be settled on a case-by-case basis. All we want to argue here is that, unlike in testing for clinical care, testing as part of public health surveillance *could* in principle be compelled, and that the differences between these circumstances make a moral difference on this issue of compelling diagnostic testing. Further, there is more going on here than just a trade-off between patient autonomy and prevention of harms to others. Adjudicating whether testing can be made mandatory requires considering issues about how data is used and whether there are ancillary obligations owed to test subjects—or in short, requires considering diagnostic justice. Thanks to an anonymous referee for pushing for clarification on this point.

> requires equality by default if: (a) there are not any relevant distinguishing feature between people that legitimate unequal distribution of advantages and disadvantages or (b) we do not have reliable ways of identifying and measuring the unequal claims people may have. (Lysdahl and Hoffman 2021, 21)

For our purposes, what is considered just or unjust when it comes to the ethical considerations of diagnostic testing will depend on the primary context in which the test is being used or conducted. That is, the purpose of testing in clinical settings, as we have seen, differs from the purpose of testing in the research trial setting, which in turn also differs from the purpose of testing in the public health setting, and these differences give rise to different ethical considerations. The ethical considerations and implications differ between these domains because the considerations of *why to test* as well as *whom to test* differ.

The answer to the *why* and *whom* questions in the clinical setting is that tests should be performed on symptomatic patients in whom the test result would be likely to change the course of their clinical care (in terms of either treatment or supportive measures). If tests are scarce, however, and there are not enough such that all symptomatic patients can receive one, then distribution decisions should be made as fairly as possible. In the context of a research trial, on the other hand, the demands of diagnostic justice differ: testing should be conducted only on symptomatic patients in this context when it is *not known* whether or not the test results would change the course of their clinical care in any significant way.[3]

Finally, in the context of public health, the answer to the *why* and *whom* to test questions is that the goal of testing is to contain the disease and testing should therefore be performed as widely, and on as many individuals, as possible (or at least, as is necessary for mitigation or successful surveillance). Further, the idea behind requiring testing in this context is that it would further the goal of mitigation or containment measures: the more people who are tested, the more likely it is that the disease will be successfully contained, especially if those in the population who test positive for active infection can be effectively isolated from others. This

[3] This epistemic requirement that it not be known ahead of time whether or not the treatment is effective is known as the *principle of equipoise* (Freedman 1987). According to Freedman, equipoise is the state of genuine uncertainty within the expert medical community on the best treatment for a condition. Thus it is a state that exists when some physicians or researchers favor one treatment (or expect it to work) while others favor another (or do not expect the one being tested to work). The idea is that this epistemic principle should be adhered to because if it is already known prior to the trial that the treatment works, then running the trial is a waste of time and financial resources, while, on the other hand, if it is already known prior to the trial that the treatment does not work, then the trial participants will be put at potential risk for no reason.

raises different distribution and allocation questions than in the case of clinical uses of testing for treatment. By way of partial analogy, in the context of justice in *treatment* allocation, in general, there are few restrictions on a competent adult patient's right to refuse a treatment measure or intervention (Flanigan 2017), although there might be restrictions on a patient's right to request these things. However, this is not as clearly the case when it comes to diagnostic testing for active surveillance purposes. In this situation, diagnostic testing is conducted not (solely) for the benefit of the individual being tested, but also to protect others in the society of which the infected person is a part.[4]

Thus the answer to the question of whether it is sometimes, always, or never acceptable to force individuals to be tested in the public health context will depend on how one settles distributive questions about the burdens of testing *when it comes to containment/mitigation measures specifically*. In considering how testing resources are allocated and how the burdens and benefits of testing are distributed, the concept of diagnostic justice provides a lens through which to evaluate how these tensions can be resolved and how the different moral demands on testing can be balanced. For example, imagine that you (unfortunately) find yourself in the emergency department of your local hospital with a diagnosis of sepsis. The treatment for this condition is intravenous antibiotic therapy, generally with two or three agents (Schmidt and Mandel 2020). But suppose that the attending physician in this case decides not to treat you because she is aware that the more often any given antibiotic is prescribed, the more likely it is that bacteria in the community will develop resistance to it. So, she decides not to treat you in order to preserve the antibiotics' effectiveness (Kennedy 2021). We might or might not agree with this physician's decision, however, what we can agree on is that she is, in the process of making this decision, weighing the benefit of the intervention to the individual vs. the risk of the intervention to society at large. That is, what she is doing is weighing in on what is the most just all-things-considered action to take in the situation. This is the sort of normative reasoning that is also required when making testing/diagnostic decisions in the clinical, research and public health settings. And, in our view, this reasoning can be facilitated by taking into consideration the principle of diagnostic justice.

[4] This is similar to the situation with vaccination—which is done not just for the benefit of the individual, but also for the benefit of the society in which that person resides.

4. Two Outstanding Difficulties in COVID Testing

Testing for COVID-19 that is part of active surveillance and mitigation efforts, as well as screening for the disease to inform quarantine decisions or travel restrictions, raises two difficulties when it comes to diagnostic justice. These difficulties are outstanding, in the sense that they have not been adequately addressed in testing policy and thus different kinds of COVID-19 testing policies may fail to meet the demands of diagnostic justice. Though testing for COVID-19 as part of the response to the pandemic was put together on the fly in the face of the global health emergency posed by the disease, it is important to understand these difficulties so as to fine tune testing policy for future public health emergencies.

A Diagnostic Misconception?[5]

A central tenet of the ethics of clinical research since the Belmont Report has been the separation of therapy from research (Emanuel et al. 2000). Revelations about the deeply unethical Tuskegee Syphilis studies in the United States showed that blurring boundaries between research and therapy can cause enormous difficulties, making exploitation of subjects much easier and complicating the exercise of an individual's right to withdraw from an experiment, among other issues.[6] It is generally accepted that, in order for clinical research to be ethical, therapy must be detached from research, in practice and in the understanding of research subjects.

Public health surveillance is similarly detached from therapy, in that the goals of public health surveillance are different from the goals of individual patient therapy. However, as happens in clinical research, individuals may not understand this difference. Patients' participation in research because they mistake it for therapy is known as the *therapeutic misconception* (Applebaum et al. 1987; Miller and Rosenstein 2003). The therapeutic misconception raises significant problems for clinical research; it may compromise informed consent, particularly in cases where participants may believe that participation in the trial is actually tantamount to a novel form of treatment, when in fact they may be assigned to a control group

[5] We owe Peter Jaworski for suggesting this term to us.

[6] It is necessary to note that a complicating factor in this case is the deep and abiding systemic racism present in the United States, which shaped the Tuskegee case and was responsible for so many of its features. The issue in Tuskegee was not just that there was a blurring of the researcher/clinician roles, it was that Black individuals were preyed upon and treated as research materials in the guise of providing them with "care".

and may receive little to no (medical) benefit from the trial at all.[7] How to deal with the therapeutic misconception in clinical trials has been a significant subject of debate (Applebaum et al. 1987).

Something very much like the therapeutic misconception may be operating in instances of disease surveillance as well. Individuals who consent to testing may not fully understand how their testing data will be used by public health decision-makers, may not understand procedures such as the deidentification of data or its use in contact tracing, and may believe that by submitting to testing, they will be facilitating their own clinical care. As an analogy, consider a study of adults in the UK about their attitudes towards contact tracing via smartphone (Williams et al. 2021). In this study researchers found that misconceptions about contact tracing data were widespread; individuals believed that contact tracing data would allow others to identify themselves, believed that contact tracing data had a kind of diagnostic function (to identify close contacts with COVID-19 so that they could understand their own risk of exposure), and did not understand how the data was being used by the government. What attitudes individuals have towards testing is an empirical question, and no doubt there will be significant research on this in the future; but it is not hard to imagine that similar misconceptions are involved with COVID-19 testing, at least at the present time.

This poses a difficulty relating to diagnostic justice for three reasons. First, individuals may be submitting to testing based on mistaken understandings of the use of the data and the purpose of the testing. As in the case of the therapeutic misconception in research ethics, this may compromise individuals' ability to give informed consent. Second, these misconceptions may be playing a part in motivating participation in testing in ways that raises worries about exploitation. In countries such as the United States where testing has been voluntary, it is possible that beliefs about the clinical relevance of testing data have played a part in individuals submitting to testing. And third, the opposite may be occurring—misconceptions about testing may play a part in keeping some individuals from submitting to testing at all, thus complicating the active surveillance measures necessary to mitigate the pandemic.

Added together, this raises a question about whether testing policy is exploiting these misconceptions to gather data. If that is the case, then testing policy, in order to be effective for active surveillance, would be

[7] They may be benefited in that they identify with the goals of the trial, and so even if participation doesn't impact their health, they may consider it a benefit to have helped further the goals of the trial. Hans Jonas famously argued that identification with the goals of a clinical trial in this strong sense was a necessary condition for a clinical trial to be morally acceptable (Jonas 1969).

depending on a widespread diagnostic misconception—to perform active surveillance, testing policy is intentionally leaving a fuzzy line between clinical and public health uses of testing, and depending on the fuzziness of the situation to leave a gap in which individuals are motivated to seek testing under mistaken pretenses. This is an issue of diagnostic justice because it raises a major concern about fairness—if individuals are seeking testing because they believe it is part of getting care, and yet it neither furthers their own care goals nor is necessary for individual care, individuals are taking on the burden (however minimal that burden is) of testing without any benefit.[8]

As with some forms of clinical research, testing for COVID-19 surveillance also involves blurred lines between the collecting of data for research and the collecting of data for therapeutic purposes. Ideally, these two domains, along with their differing aims and ethical considerations should be kept separate. However, during public health emergencies, these lines are almost necessarily blurred. Clinicians become researchers and *vice versa* and are suddenly tasked with the considerations of both knowledge acquisition and patient care. We have seen this in the current pandemic, as data gathered in the course of the clinical care of COVID-19 patients has both been made public and has been used to inform public health decision-making. For example, testing data from clusters identified at the beginning of the pandemic were instrumental in establishing that the disease is spread via aerosol transmission (Hamner et al. 2020). Unlike in (well-designed) clinical trials, there are no clear protocols on how to keep these roles separate. Further, this blurring of clinical and public health surveillance roles for testing and data gathering, both in the understanding of individuals submitting to testing and in the practices of both clinicians and researchers, could pose significant problems in the future. This is an area that requires further investigation and would greatly benefit from the development of clear protocols.

Use of Data and Impacts on Communities

It is well recognized that participation in research does not always benefit the individual participants involved, and because of this, what benefits are owed to research subjects has itself been a subject of intense debate within the ethics of clinical research (Richardson 2012).

Similarly, participation in active surveillance by submitting to testing does not always benefit individuals or even their communities, and in fact can be used to inform decision-making that could potentially *harm* these

[8] Thanks to an anonymous referee, for pushing us to clarify this point.

communities. One of the major features of the COVID-19 pandemic has been the significant disparities in morbidity/mortality rates among different communities, with Hispanic, Latinx, Black, Indigenous, and Pacific Islander populations disproportionately affected by the disease (Hooper et al. 2020). These dynamics were noticed very early on in the pandemic, and yet data gathered from surveillance has done little to make a dent in this disparity. This is a significant concern for diagnostic justice; if testing as part of active surveillance reveals such significant and morally arbitrary disparities, it should, ideally, also inform policies that address these problems. Yet in the case of COVID-19, the opposite has been the case; upticks in infections revealed by active surveillance testing informed policies that seemed to have little to no impact on these disparities. A vivid example of this has been the US state of California, where an early lockdown likely mitigated the impact of the pandemic in the early months of the pandemic (Friedson et al. 2021), but where there have been massive disparities between lower-income and higher-income communities and white and Latinx communities in their respective burdens of COVID-19 morbidity and mortality (Hsu and Hayes-Bautista 2021). Why data revealed from active surveillance indicated these disparities but policy did not adjust accordingly is a major issue that must be addressed in the wake of the pandemic. If active surveillance reveals such a disparity, but policy does nothing to ameliorate it, this looks like a significant failure of diagnostic justice, as the public health purposes of testing and compliance with testing requirements by community members did not result in any action that ameliorated the effects of the pandemic.

The primary function of data gathered from active surveillance has, so far, been to inform when to impose different restrictions on businesses, schools, and other public activities. Different communities have experimented with various metrics in an effort to determine when it is safe to permit school openings, religious services, dine-in service at restaurants, and the like. As an example, New York City, in the United States, established fairly early on in the pandemic a metric of a 3% test positivity rate for opening public schools (Shapiro 2020). These restrictions, however, do not benefit or harm everyone equally; in New York City, the effects of closing public schools have primarily been felt by lower-income communities (Agostinelli et al. 2020). There are also worries about the disproportionate long-term effects of lockdowns from lost income, mental health impacts, and the like (Winsberg et al. 2020).[9] During the COVID-19 pandemic, testing data has informed these policies. Testing data, then,

[9] We bracket here any comment on Winsberg et al.'s claim that these long-term effects show that trade-offs from lockdowns raise a high epistemic barrier to imposing such lockdowns, and that this barrier was not met in the early months of the pandemic (Winsberg et al. 2020).

can be used in such a way that informs policy-decisions that impose burdens, but in which burdens are not distributed equitably, in which burdens fall disproportionately on some communities and not others. If testing data gathered during active surveillance informs policies that not only do not ameliorate the impacts of the pandemic on disproportionately affected communities, but actually generate some significant harms of their own, then this also looks like a significant failure of diagnostic justice.

5. Implications

Our discussion of diagnostic testing and diagnostic justice has implications not just for COVID-19 testing but for testing policy for future public health emergencies. As we have seen, testing for COVID-19 as part of active surveillance efforts can involve a blurring of the boundaries between public health and clinical medicine. Since test results are obviously relevant for an individual's health, testing as part of active surveillance and mitigation efforts at least has some relevance for individuals, even if that is not the primary goal of the testing. Given this, it may be that testers have obligations to individuals who report for testing as part of active surveillance efforts, even if the primary aim is not clinical but is to provide data for mitigation efforts. These obligations, for testing as part of active surveillance, may be minimal: timely return of results, clinical advice and direction to care resources, communication of results to individuals in a clear fashion, and the like may be sufficient to discharge the duties resulting from the partial entrustment of individuals' health to testers. However minimal, meeting these requirements may be necessary to ensure that benefits from testing are distributed equitably. Some individuals may be better placed to take advantage of information gained from testing without additional resources or aid from public health officials. Building in resources to meet obligations of care to those who submit to testing may be necessary to help remove these inequities, and ensure that those who submit to testing receive some (clinical) benefit from doing so, as well as those who benefit from mitigation efforts.

Though minimal, this hasn't always been the case with active surveillance measures during epidemics. During the 2013-2016 Upper West Africa Ebola epidemic, the focus throughout, from the very earliest days, was on containment, instead of care (Farmer 2020). Pressure from the world community on Guinea, Sierra Leone, and Liberia led to a channeling of resources into identification and isolation of cases, in the hopes of breaking transmission chains, and this extended as well to testing and contact tracing. Much of the containment and mitigation effort was put in the hands of the military, which employed coercive measures aimed at containment

(such as the infamous *cordon sanitaire*) (McNeill 2014). As the medical historian Frank Snowden argues, the response to Ebola involved a resurrection of the tactics used to fight infectious disease in the dark ages of medicine, rather than a 21st century, biomedically sophisticated effort aimed at both care and mitigation:

> Many of the coercive means adopted echoed early modern Europe's effort to defend itself against bubonic plague (…). Compulsory treatment facilities surrounded by troops even closely resembled lazarettos. Daniel Defoe would have found the response familiar. (Snowden 2019, 495).

Besides the obvious wrong of failing to provide even minimal supportive care to those suffering from Ebola Virus Disease, this also hampered mitigation efforts, as the (correct) perception that public health authorities (including some, but not all, foreign support) were more interested in containment than in caring for the sick sowed distrust and resentment, and led to (sometimes violent) backlash among the population of the three most affected countries. Though testing during the Upper West African Ebola epidemic was not nearly on the scale of the current worldwide efforts to test for SARS-CoV-2, and there are many relevant differences in the dynamics of the two epidemics, the contrast between the two events shows how employing active surveillance without providing any clinical support leads not just to serious harms but is counterproductive to mitigation.[10] This has important implications for global health ethics and public health policy looking forward: the separation of care from mitigation is neither normatively nor practically possible, and active surveillance measures, including testing for this purpose, must recognize the requirements of care to the individuals being tested in order to equitably distribute the burdens and benefits of testing, even if the primary goals of surveillance are not clinical.

6. Conclusion

We have argued in this paper that considerations of diagnostic justice generate moral demands on testing policy as part of public health

[10] There are many reasons, of course, for the differences between the two events: the Upper West Africa Ebola epidemic occurred in a region with minimal clinical resources (Farmer 2020), the epidemic was concentrated in Upper West Africa despite some sporadic imported infections (and limited secondary transmission) elsewhere in Africa, Europe, and the United States, and the different stigmas, biases, and prejudices about Ebola and those suffering most from it during the epidemic made it far easier to "other" those in need of care and thus to direct resources elsewhere than has been the case with COVID-19, although there is also plenty of stereotyping of individuals susceptible to the disease in the latter case as well (Aronson 2020).

surveillance during infectious disease epidemics. The current and ongoing SARS-CoV-2/COVID-19 pandemic has revealed many of the dynamics involved with testing as part of active surveillance during these events and provided important lessons for the general question of what would constitute an ethical testing regime for active surveillance during epidemics. This, unfortunately, looks likely to be a significant question for global health in the foreseeable future. The first two decades of the 21st century have already seen a number of significant public health events involving novel and emerging pathogens—SARS, H1N1, Ebola, and now COVID-19. Collectively, these have already cost the lives of millions of people, in the form of premature death from infection and illness. There are plenty of reasons to believe this is not just bad luck; some of the dynamics of our world—further encroachment into the wildland-urban interface (which provides increased opportunities for zoonosis), intensifying urbanization of the world's population, the high volume of international air travel, and continuing, morally pernicious disparities in access to basic health care resources in many parts of the world—all provide ample opportunities for emerging pathogens to spark epidemics (Bollyky 2018).[11] A just and sustainable world will require just and sustainable global health policy, which includes testing protocols for public health surveillance that meet the demands of diagnostic justice.

Acknowledgments

Versions of this paper were presented at the Georgetown Institute for the Study of Markets and Ethics at Georgetown University in April 2021, and the Philosophical Perspectives on COVID-19 Workshop at the University of Johannesburg in May 2021. Thanks to audiences at both, for their helpful comments and suggestions (and thanks to Sahar Akhtar for inviting us to Georgetown and to Alex Broadbent for organizing the Johannesburg workshop). The idea for this paper came out of an exchange between both authors during the Q&A for a presentation one author (Kennedy) gave to the PDXPhiSciNOW philosophy of the life sciences workshop in December 2020. Thanks to Mark Bedau for organizing the workshop.

[11] The causal claims involved in theories about the vulnerability of the contemporary world to infectious disease generate interesting questions in the philosophy of science in their own right; but some of the narratives and rhetoric in the presentation of these claims can echo problematic ideas about developing countries from past decades. Some of this is the case with Bollyky's treatment, especially his discussion of the role of urbanization in developing countries and population increases due to progress in combating childhood mortality (Bollyky 2018). Others draw different lessons; Deaton (2013) and Farmer (2020), for instance, see the unique zoonotic opportunities provided by urbanization and encroachment on the urban-wildland interface in developing countries as evidence of the severe risks and injustices posed by lack of public health infrastructure and clinical resources; or rather, as evidence not that, as Bollyky puts it, "the world is getting healthier in worrisome ways", but rather that persistent injustices in access to health care and other basic goods create significant risks for all.

REFERENCES

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti. 2020. "When the Great Equalizer Shuts down: Schools, Peers, and Parents in Pandemic Times." w28264. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w28264.

Appelbaum, P. S., L. H. Roth, C. W. Lidz, P. Benson, and W. Winslade. 1987. "False Hopes and Best Data: Consent to Research and the Therapeutic Misconception." *The Hastings Center Report* 17 (2): 20–24.

Aronson, Louise. 2020. "Age, Complexity, and Crisis — a Prescription for Progress in Pandemic." *New England Journal of Medicine* 383 (1): 4–6. https://doi.org/10.1056/NEJMp2006115.

Beauchamp, Tom L., and James F. Childress. 2020. *Principles of Biomedical Ethics*. 8th ed. New York: Oxford University Press.

Begley, Sharon. 2020. "Who Is Getting Sick? A Breakdown of Coronavirus Risk by Demographic Factors." *STAT* (blog). March 3, 2020. https://www.statnews.com/2020/03/03/who-is-getting-sick-and-how-sick-a-breakdown-of-coronavirus-risk-by-demographic-factors/.

Berkley, Seth. 2003. "Thorny Issues in the Ethics of AIDS Vaccine Trials." *Lancet* 362 (9388): 992. https://doi.org/10.1016/S0140-6736(03)14371-1.

Bollyky, Thomas J. 2018. *Plagues and the Paradox of Progress: Why the World Is Getting Healthier in Worrisome Ways*. Cambridge, MA: MIT Press.

Center for Disease Control and Prevention. 2020. "Testing Strategies for SARS-CoV-2." Center for Disease Control and Prevention. February 11, 2020. https://www.cdc.gov/coronavirus/2019-ncov/lab/resources/sars-cov2-testing-strategies.html.

Cheng, Matthew P., Jesse Papenburg, Michaël Desjardins, Sanjat Kanjilal, Caroline Quach, Michael Libman, Sabine Dittrich, and Cedric P. Yansouni. 2020. "Diagnostic Testing for Severe Acute Respiratory Syndrome–Related Coronavirus-2." *Annals of Internal Medicine*, M20-1301. https://doi.org/10.7326/M20-1301.

Deaton, Angus. 2013. *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton: Princeton University Press.

Emanuel, E. J., D. Wendler, and C. Grady. 2000. "What Makes Clinical Research Ethical?" *JAMA* 283 (20): 2701–11. https://doi.org/10.1001/jama.283.20.2701.

Emanuel, Ezekiel J., Govind Persad, Ross Upshur, Beatriz Thome, Michael Parker, Aaron Glickman, Cathy Zhang, Connor Boyle, Maxwell Smith, and James P. Phillips. 2020. "Fair Allocation of Scarce Medical Resources in the Time of Covid-19." *New England Journal of Medicine* 382 (21): 2049–55. https://doi.org/10.1056/NEJMsb2005114.

Farmer, Paul. 2020. *Fevers, Feuds, and Diamonds: Ebola and the Ravages of History*. New York: Farrar, Straus and Giroux.

Flanigan, Jessica. 2014. "A Defense of Compulsory Vaccination." *HEC Forum: An Interdisciplinary Journal on Hospitals' Ethical and Legal Issues* 26 (1): 5–25. https://doi.org/10.1007/s10730-013-9221-5.

———. 2017. *Pharmaceutical Freedom: Why Patients Have a Right to Self-Medicate*. New York: Oxford University Press.

Freedman, B. 1987. "Equipoise and the Ethics of Clinical Research." *The New England Journal of Medicine* 317 (3): 141–45. https://doi.org/10.1056/NEJM198707163170304.

Friedson, Andrew I., Drew McNichols, Joseph J. Sabia, and Dhaval Dave. 2021. "Shelter-in-place Orders and Public Health: Evidence from California during the Covid-19 Pandemic." *Journal of Policy Analysis and Management* 40 (1): 258–83. https://doi.org/10.1002/pam.22267.

Furukawa, Nathan W., John T. Brooks, and Jeremy Sobel. 2020. "Evidence Supporting Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 While Presymptomatic or Asymptomatic - Volume 26, Number 7—July 2020 - Emerging Infectious Diseases Journal—Cdc." Accessed November 9, 2021. https://doi.org/10.3201/eid2607.201595.

Hamner, Lea, Polly Dubbel, Ian Capron, Andy Ross, Amber Jordan, Jaxon Lee, Joanne Lynn, et al. 2020. "High Sars-CoV-2 Attack Rate Following Exposure at a Choir Practice—Skagit County, Washington, March 2020." *MMWR. Morbidity and Mortality Weekly Report* 69 (19): 606–10. https://doi.org/10.15585/mmwr.mm6919e6.

Hsu, Paul, and David E. Hayes-Bautista. 2021. "The Epidemiology of Diversity: Covid-19 Case Rate Patterns in California." *Journal of Immigrant and Minority Health* 23 (4): 857–62. https://doi.org/10.1007/s10903-021-01159-x.

Jonas, Hans. 1969. "Philosophical Reflections on Experimenting with Human Subjects." *Daedalus* 98 (2): 219–47.

Jr, Donald G. McNeil. 2014. "Using a Tactic Unseen in a Century, Countries Cordon off Ebola-Racked Areas." *The New York Times*, 2014, sec. Science.

https://www.nytimes.com/2014/08/13/science/using-a-tactic-unseen-in-a-century-countries-cordon-off-ebola-racked-areas.html.

Karlsson, Annika C., Marion Humbert, and Marcus Buggert. 2020. "The Known Unknowns of T Cell Immunity to COVID-19." *Science Immunology* 5 (53): eabe8063. https://doi.org/10.1126/sciimmunol.abe8063.

Larijani, Mona Sadat, Amitis Ramezani, and Seyed Mehdi Sadat. 2019. "Updated Studies on the Development of HIV Therapeutic Vaccine." *Current HIV Research* 17 (2): 75–84. https://doi.org/10.2174/1570162X17666190618160608.

Lasagna, L. 1968. "Some Ethical Problems in Clinical Investigation." *South African Medical Journal* 42 (1): 2–5.

Lysdahl, Kristin Bakke, and Bjørn Hofmann. 2020. "Overutilization of Imaging Tests and Healthcare Fairness." In *Philosophy of Advanced Medical Imaging*, edited by Elisabetta Lalumera and Stefano Fanti, 99–111. SpringerBriefs in Ethics. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-61412-6_8.

Miller, Franklin G., and Donald L. Rosenstein. 2003. "The Therapeutic Orientation to Clinical Trials." *New England Journal of Medicine* 348 (14): 1383–86. https://doi.org/10.1056/NEJMsb030228.

Oleske, Denise M. 2010. "Screening and Surveillance for Promoting Population Health." In *Epidemiology and the Delivery of Health Care Services: Methods and Applications*, 131–50. Boston, MA: Springer. https://doi.org/10.1007/978-1-4419-0164-4_5.

Onder, Graziano, Giovanni Rezza, and Silvio Brusaferro. 2020. "Case-Fatality Rate and Characteristics of Patients Dying in Relation to Covid-19 in Italy." *JAMA*. https://doi.org/10.1001/jama.2020.4683.

Paltiel, A. David, Amy Zheng, and Rochelle P. Walensky. 2020. "Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United States." *JAMA Network Open* 3 (7): e2016818. https://doi.org/10.1001/jamanetworkopen.2020.16818.

Richardson, Henry S. 2007. "Gradations of Researchers' Obligation to Provide Ancillary Care for HIV/AIDS in Developing Countries." *American Journal of Public Health* 97 (11): 1956–61. https://doi.org/10.2105/AJPH.2006.093658.

———. 2012. *Moral Entanglements: The Ancillary-Care Obligations of Medical Researchers*. Oxford: Oxford University Press.

Shapiro, Eliza. 2020. "New York City to Close Public Schools Again as Virus Cases Rise." *The New York Times*, 2020, sec. New York.

https://www.nytimes.com/2020/11/18/nyregion/nyc-schools-covid.html.

Sharfstein, Joshua M., Scott J. Becker, and Michelle M. Mello. 2020. "Diagnostic Testing for the Novel Coronavirus." *JAMA* 323 (15): 1437–38. https://doi.org/10.1001/jama.2020.3864.

Snowden, Frank M. 2019. *Epidemics and Society: From the Black Death to the Present*. Open Yale Courses Series. New Haven: Yale University Press.

Spellberg, Brad, Travis B. Nielsen, and Arturo Casadevall. 2021. "Antibodies, Immunity, and COVID-19." *JAMA Internal Medicine* 181 (4): 460. https://doi.org/10.1001/jamainternmed.2020.7986.

Truog, Robert D., Christine Mitchell, and George Q. Daley. 2020. "The Toughest Triage — Allocating Ventilators in a Pandemic." *New England Journal of Medicine* 382 (21): 1973–75. https://doi.org/10.1056/NEJMp2005689.

Webb Hooper, Monica, Anna María Nápoles, and Eliseo J. Pérez-Stable. 2020. "Covid-19 and Racial/Ethnic Disparities." *JAMA* 323 (24): 2466. https://doi.org/10.1001/jama.2020.8598.

Williams, Simon N., Christopher J. Armitage, Tova Tampe, and Kimberly Dienes. 2021. "Public Attitudes towards COVID-19 Contact Tracing Apps: An UK-based Focus Group Study." *Health Expectations* 24 (2): 377–85. https://doi.org/10.1111/hex.13179.

Winsberg, Eric, Jason Brennan, and Chris W. Surprenant. 2020. "How Government Leaders Violated Their Epistemic Duties during the SARS-CoV-2 Crisis." *Kennedy Institute of Ethics Journal* 30 (3): 215–42. https://doi.org/10.1353/ken.2020.0013.

European Journal of Analytic Philosophy EuJAP | Vol. 17 | No. 2 | 2021
https://doi.org/10.31820/ejap.17.3.2

ADAPT TO TRANSLATE – ADAPTIVE CLINICAL TRIALS AND BIOMEDICAL INNOVATION

Daria Jadreškić[1]

[1] University of Klagenfurt

Original scientific article – Received: 16/05/2021 Accepted: 26/09/2021

ABSTRACT

The article presents the advantages and limitations of adaptive clinical trials for assessing the effectiveness of medical interventions and specifies the conditions that contributed to their development and implementation in clinical practice. I advance two arguments by discussing different cases of adaptive trials. The normative argument is that responsible adaptation should be taken seriously as a new way of doing clinical research insofar as a valid justification, sufficient understanding, and adequate operational conditions are provided. The second argument is historical. The development of adaptive trials can be related to lessons learned from research in cases of urgency and to the decades-long efforts to end the productivity crisis of pharmaceutical research, which led to the emergence of translational, personalized, and, recently, precision medicine movements.

Keywords: *adaptive clinical trials; randomized controlled trials; reliability; urgency; precision medicine; translational medicine; the productivity crisis*

Correspondence: daria.jadreskic@aau.at

1. Introduction

Adaptive clinical trials have been at the forefront of the efforts to mitigate the ongoing coronavirus pandemic due to their shorter duration and flexible design, which allows for accelerated assessment and the timely implementation of new vaccines and therapies (WHO 2020; Stallard et al. 2020; Branch-Elliman, Elwy, and Monach 2020; London and Kimmelman 2020). Adaptive trials are a subset of randomized controlled trials (RCTs), in which one or more features of the design can be changed during the trial's course based on interim results from the data accumulated early on.[1] Although they use control groups and randomization of patients to either the experimental or the control treatment, they differ from the standard RCTs by the absence of a fixed design. A fixed trial is first designed, conducted, and then analyzed upon completion, with no intermediate steps. In cases in which quick action is needed and standard RCT evidence is not available and takes too long to acquire, observational and other types of evidence need to provide temporary guidance. Adaptive design trials enable this by generating results based on observing patient responses and conducting interim analyses, in this way integrating evidence from experimentation with observational evidence and preclinical data.

Recently, London and Kimmelman have argued for the usage of multi-arm and seamless adaptive design trials, stating that "one lesson of the current outbreak is that expeditious research in a crisis situation is feasible" (2020, 477). If responsible expeditious research via adaptive design is feasible, should its methodology be used more widely, also in non-crisis contexts? To what extent are adaptive trials a valid, or even superior alternative to fixed RCTs in clinical research? If yes, on which grounds and under what circumstances? A conjoined ethical and epistemological discussion is in place. The aim of this paper is twofold: to outline some of the advantages and limitations of adaptive trials, and to specify the conditions that contributed to their development and implementation in clinical practice. This will make a case for their usage, but not in all contexts.

The first argument advanced in this paper is normative: responsible adaptation should be taken seriously as a new way of doing clinical research, but only insofar as a valid justification, sufficient understanding, and adequate operational conditions for the introduction of adaptive measures are provided. The most common obstacles to their implementation are local and practical, rather than general and principled. The greatest danger to the integrity of clinical research is shared across

[1] There can be non-randomized and uncontrolled trials, including adaptive trials, but they do not satisfy regulatory standards and their limitations are well documented.

different designs: it is, on the one hand, the ineliminable uncertainty of experimenting, and on the other, it is the intrusion of unwanted bias, such as sponsorship bias, or more broadly, preference bias (Wilholt 2009). However, both dangers hold for fixed and adaptive trials alike, and should not downplay positive aspects of adaptation.

The second argument is historical: the presence of adaptive trials as one of the potential drivers of biomedical innovation can be related not only to lessons learned from research in cases of urgency, but also to the decades-long efforts to end the productivity crisis of pharmaceutical research, which led to the emergence of translational, personalized, and more recently, precision medicine movements. These efforts have motivated new methods, organization, and relations between research stakeholders. Biomedical innovation has been spurred by investments in education and training in translational research, promotion of interdisciplinarity, collection of a variety of data- and bio-banks, developments in bioinformatics, calls for inclusion of patients in healthcare decision-making, and a general focus on the (re)organization of basic-clinical research interface via private-public partnerships. This has contributed to a broadening of clinical research teams to include experts in bioinformatics, statistics, and other big data skills which have enabled, among else, innovations in clinical trial design.

The ratio of randomization to different treatment arms in adaptive trials may not be equal or consistent throughout the trial's course, so the term 'adaptive' sometimes primarily characterizes randomization, such as in "outcome-adaptive randomization" (Berry 2011). Other adaptations include changes in sample size, treatment dose, or patient allocation ratio (Pallmann et al. 2018, 2). Adaptation can also mean abandoning treatment arms, stopping the trial early because of evident success or a lack of efficacy, or identifying and recruiting patients who are most likely to benefit from the treatment. Adaptive trials can assess several treatments in a single trial, or seamlessly merge different trial phases into only one trial. Adaptations need to be preplanned and modeled before the onset of the trial to preserve its integrity and generate valid results (Pallmann et al. 2018, 10-11). Without planning, rigorous execution and analysis, there is an increased risk of introducing bias into the trial. Results can be difficult to interpret due to a higher tolerance for false positives, in other words, for cases of observed beneficial effects whose cause is wrongly attributed to the experimental treatment.

A departure from the fixed RCT standard predates the coronavirus pandemic. Adaptive trials have been used both in urgent circumstances such as the 2013-2016 Ebola virus (Henao-Restrepo et al. 2017; Calain

2018) and earlier the AIDS epidemic (Epstein 1996), but also for evaluating therapies in the domain of precision medicine. If the mechanism of the experimental intervention is well understood, for example, because of the possibility to match therapies with subgroups of patients based on genomic data, the trial can be designed to recruit only patients who will benefit from the treatment. Adaptive trials are thus being increasingly used for evaluating the efficacy of cancer therapies and other targeted interventions (Riley 2016; Garralda et al. 2019), and both EMA and FDA have included them in their regulatory schemes (EMA 2017; FDA 2019).

I section 2, I discuss two cases of adaptive trials: the azidothymidine (AZT) trial in the 1980s and Ebola ca Suffit! trial in 2015. These two trials present milestones for the usage of adaptation in the context of crisis. Motivations for conducting adaptive trials are identified, as well as the trade-offs permeating the decision to rely on them. Section 3 puts forward the bulk of the normative argument. I draw on London and Kimmelmann's (2020) lessons from the ongoing coronavirus pandemic to show that reliable adaptation is alive and well and that the tension between reliability and speed in clinical research can be dissolved, but only under adequate operational conditions for running large-scale, multi-arm adaptive trials. I use the notion of operational exceptionalism to depict the current situation in which adaptive trials can be successfully implemented only via "carefully orchestrated protocols" (London and Kimmelmann 2020, 477) in big research centers with close ties to industry and policy makers. In section 4, I present a cluster of adaptive measures developed as part of clinical research in precision medicine. New conditions under which adaptations can be preferred to fixed RCTs are identified. In section 5, the historical path to precision medicine is outlined. The focus is on the emergence of different biomedical initiatives in the big data era that have brought new ways of generating and assessing evidence, together with innovations in clinical research which are following up on the advances.

The concluding section sums up the two arguments. Since the material, infrastructural, computational, and organizational conditions for conducting adaptive trials are at hand more than ever before, the case for their wider usage is made stronger. Still, there are practical and logistical drawbacks to the possibility of successfully implementing complex interventions such as adaptive trials across the board. Their recent successful uptake in assessing Covid-19 vaccines and treatments gives us much reason for optimism, but almost as much for caution. Adaptation should not mean that anything goes, but rather that everything is in place to make a balanced judgment based on available evidence and cooperative engagement of various interested parties. Inevitably, these hard choices are made in face of great uncertainty and nested interests.

2. Adaptive Trials in Epidemics

In this section, I present two cases of adaptive trials conducted in the urgent context of an ongoing epidemic. In these cases adaptation was chosen as a consequence of exceptional circumstances, prompted by ethical reasons to balance potential harms in a particular way.

The first case is the controversial AZT trial during the AIDS epidemic in the late 1980s, known for the groundbreaking role played by patient advocacy and citizen science (Epstein 1996). The first drug for AIDS, azidothymidine (AZT), was approved more quickly than subsequent therapies, in part because of the pressure for quick approvals coming from patients' advocacy groups and the fact that there was no efficient therapy available. Although planned as a fixed, double-blinded, randomized, placebo-controlled trial, control groups were eventually excised from the trial so that more patients could get the medication immediately. This practice is considered adaptive by clinical research standards, as volunteers would normally be randomly assigned to either the treatment or the control arm equally, and the randomization ratio would be fixed until the end of the trial. Because there was no therapy for AIDS and the patients' prospects were poor, many of them felt that they had nothing to lose. Potential harms associated with accelerated access to the experimental therapy were considered acceptable for many patients seeking help. In a record time, AZT was approved in 1987 after it had shown beneficial effects. However, the drug was not as successful as it was first thought. A three year follow up study of its effectiveness conducted on two thousand patients showed that patients in the placebo group were more likely to survive the three years of study than patients on AZT and that the drug had serious side effects and almost no benefits after a certain period of usage (Crewe 2018). It was later shown that AZT has beneficial effects, but only in combination with other medications, which is how it is still being prescribed and used.

The AZT trial is controversial to date. Should the drug have been approved? At the time, patients were pressuring the FDA for quicker approval and the FDA responded by adjusting the standards to meet their requests. This was done without much understanding of either the virus, the intervention, or the alternative trial design. There was no concept of an 'adaptive trial' at this stage—the trial was planned with a fixed design, only eventually accelerated, and adapted on the go. Concerns about patient recruitment and management strategies have been raised, such as the lack of coordination across twelve research centers that participated in the trial (Sonnabend 2011). There was a striking difference in mortality between the treatment and the control group (1 to 19 in the first 120 days) which decided in favor of expanding the treatment arm, but according to

Sonnabend, this discrepancy might have been an effect of biased patient selection and management. He also reports that the dose of initially administered AZT has been criticized for being too high. This might have led to beneficial short-term effects, but damaging long-term effects. Additionally, suspicions were raised about the practical limitations to blinding in such a study: The drug causes changes in routine blood counts that investigators need to see. Therefore we must conclude that investigators could know who was receiving AZT or placebo (Sonnabend 2011).

Doubts about the first AZT trial are primarily related to preference bias. Preference bias

> occurs when a research result unduly reflects the researchers' preference for it over other possible results. (…) It works (…) by increasing the likelihood of the preferred outcome rather than by bluntly fabricating it. (Wilholt 2009, 92)

It is not clear that this is what happened in the 1987 AZT trial, but if anything worrisome had happened, it seems to fall under the scope of preference bias. However, such subtle biasing is not attached to a particular design and it, unfortunately, permeates the landscape of biomedical and especially, pharmaceutical research (see Biddle 2007). Researchers, producers, policy makers, and patients had high hopes about AZT efficacy in absence of AIDS treatments. Everyone wanted the drug to work, and the trial was exceptional in both its urgent undertaking and its striking first outcomes.

Despite possible problems with the trial, the regulators had good reasons to approve the drug in face of reported evidence. Besides, pharmacovigilance, or monitoring for side effects of the drugs on the market, is in place to identify problems that might have been missed on the scale of pre-approval research. Time-spans of drug activity, effects after prolonged usage, and usage for different subgroups of patients can differ drastically. Benefits, side-effects, and long-term effects show at different times, and risk is inevitable: between waiting for the approval too long (denying people access to potentially effective therapy) and granting the approval too quickly (allowing for the provision of ineffective or harmful therapy). The balance was struck in the AZT case on the side of quick yet possibly unreliable assessment, although promising at the time, as opposed to waiting for more evidence in face of great public outcry. The therapy was made available, followed up, and finally, restricted in use. In addition to ethical considerations about research in exceptional circumstances, the AZT trial brought to attention patients' roles as advocates and partners in

healthcare decision-making. Today we find appeals to caution when it comes to such adaptations, but also tools and skills developed to plan and simulate a trial's course should adaptive interventions be made (Pallmann et al. 2018, 10-11). Special care needs to be taken to ascertain the best dosage, optimal sample size and representativeness, and comparators to the experimental treatment. Additional staff and resources need to be in place to reconcile the need to make interim analysis with the need to keep the results blinded. Local discrepancies between research centers should be minimized by transparent protocols and centralized oversight.

The second case has attracted philosophical attention both because of ethical challenges related to responses to emergencies and disasters (Calain 2016), but also because of a conjoined ethical-epistemic interest in innovative trial design (Upshur and Fuller 2016; Varghese 2021a, 2021b). In 2015 a phase III trial called 'Ebola ça Suffit!' ('Ebola, that's enough!') was conducted for testing recombinant vesicular stomatitis virus-Zaire Ebola vaccine (rVSV–ZEBOV) against Ebola virus disease. The design of the trial was not standard, due to time constraints, a limited amount of vaccine supplies, ethical concerns regarding the adoption of research methodology, and logistics and field operational challenges (Varghese 2021a, 2021b; Calain 2018). 'Ebola ça Suffit!' was a result of collective efforts to respond to the 2013-2016 West African Ebola epidemic that had caused the death of more than 11,000 people (Calain 2018). In August 2014, the Ebola epidemic was declared a public health emergency of international concern, and the World Health Organization (WHO) set up a panel of experts to consider ethical permissibility of testing potentially effective interventions for the disease in an accelerated manner. Within a few months, novel or repurposed therapeutic agents were tested for efficacy at various locations experiencing an outbreak.

The 'Ebola ça Suffit!' ring trial used cluster randomization instead of individually controlled randomization, and a delayed vaccination arm as the control group instead of a placebo control group, to mitigate the transmission of the disease in case of evidence of efficacy. Upon confirming a case of the Ebola virus, a ring (cluster) of all infected persons' contacts was established, as well as the contacts of their contacts (Henao-Restrepo et al. 2017). The clusters were assigned to either immediate vaccination or a delayed vaccination arm, allowing both groups to receive the vaccine, as opposed to treating the control group with a placebo. The randomization stopped after four months to allow the immediate provision of the vaccine to more adults, and to include younger age groups sooner (WHO 2015). The vaccine was approved for 'compassionate use' in outbreaks, meaning that it had been proven sufficiently safe and effective to be recommended, although it had not yet been formally approved by a

full regulatory process. According to later correspondence in *The Lancet*, the efficacy estimate of the vaccine remained at 100% despite concerns about bias in the research design (Longini et al. 2018; Metzger and Vivas-Martínez 2018). The vaccine eventually contributed to the suppression of the 2013-2016 Ebola virus disease epidemic (Geisbert 2017; Calain 2018).

Upshur and Fuller (2016) draw on the lessons from Ebola trials to call for a philosophy of clinical trials, asserting that the "inherent trade-off between ethical requirements and scientific rigor" is not resolved "necessarily through insisting on validity over ethics, but rather in reaching consensus on what is at stake" (2016, 11). They characterize the successful implementation of the ring vaccination strategy as "evidence that alternative trial designs can work", although they are not based on classical randomization which conventionally grants validity and reliability to clinical research. In a similar vein, Varghese (2021a, 2021b) uses the distinction between epistemic and non-epistemic values to argue that non-epistemic values were rightfully prioritized over epistemic values in the case of 'Ebola ca Suffit!' The urgency of the intervention was prioritized over scientific understanding that a standard procedure would advance. In a situation in which it was necessary to stop the virus from spreading, cluster randomization was considered good enough and prioritized over individual randomization. It is important to note that randomization was not altogether avoided. Like in the AZT case, it was only adapted. In the AZT trial, control arms were dropped only when beneficial results after initial randomization were observed, while in 'Ebola ca Suffit!' randomization was applied to clusters as opposed to individuals. Additionally, control groups were excised only with a delay, when beneficial effects of the vaccine were observed. Adaptation thus did not replace randomization and controlling, it rather complemented them and made the trial feasible and apt given the circumstances.

3. Towards Operational Exceptionalism

In a recent article, London and Kimmelman (2020) argue against what they call pandemic research exceptionalism, according to which situations of crisis justify lowering research standards. They identify three problematic assumptions which underpin research exceptionalism. The first is that any evidence, even if flawed, is preferable to more demanding studies whose benefits show later. In other words, that evidence generated by a faster method is preferred to evidence generated by a slower method. The second is that scientific rigor conflicts with care. The third problematic assumption is that researchers and sponsors are allowed to exercise discretion over the

organization and design of research in times of crisis. These assumptions, they contend, underlie alarming practices in pandemic research.

> The proliferation of small studies that are not part of an orchestrated trajectory of development is a recipe for generating false leads that threaten to divert already scarce resources toward ineffective practices, slow the uptake of effective interventions because of an inability to reliably detect smaller but clinically meaningful benefits, and engender treatment preferences that make patients and clinicians reluctant to participate in randomized trials. (London and Kimmelman 2020, 476)

The small studies referred to in this passage are numerous clinical trials that have been flourishing after the outbreak of the coronavirus epidemic, often investigating similar hypotheses in absence of coordinated oversight, rushing to publish results based on spurious correlations, and lacking adequate power to detect clinical benefit. Importantly, they are not a part of an "orchestrated trajectory of development", in other words, of a coordinated translational enterprise. When London and Kimmelman complain about "patients and clinicians being reluctant to participate in randomized trials", it is the adaptive randomized trials they refer to, which, according to them, hold a key to upholding both the standards of research excellence and time sensitivity.

> Sponsors, research consortia, and health agencies should prioritize research approaches that test multiple interventions, foster modularity, and permit timely adaptation. (…) Adaptive designs allow flagging interventions to be dropped quickly and promising alternatives to be added with fewer delays than would be incurred from the design and approval of new studies. (London and Kimmelman 2020, 477)

The argument is that adaptive trials should be undertaken under careful coordination in big research centers with the ability to conduct and analyze them, and not that any adaptation will satisfy. Quite the contrary—adaptation is here understood as a powerful, but demanding and complex method that can only work when five conditions of informativeness and social value are met, and under strict guidance and oversight.

The conditions identified by London and Kimmelman are importance, rigorous design, analytical integrity, complete, prompt, and consistent reporting, and feasibility. The condition of importance requires that trials address evidence gaps, aiming to detect effects that are "realistic but

clinically meaningful" (London and Kimmelman 2020, 476). An example of bad practice would be to concentrate resources on identical clinical hypotheses, creating competition for recruitment, and a neglect of other hypotheses, as was the case at the time of hydroxychloroquine hype when many trials were conducted in the US to test its efficacy for alleviating Covid-19 symptoms. Rigorous design is ascertained by randomization, blinding, controlling, and using meaningful endpoints. An example of bad practice would be "to forego a dummy comparator and use a nonvalidated surrogate endpoint" (London and Kimmelman 2020, 477). Analytical integrity means that designs should be "prespecified in protocols, prospectively registered, and analyzed in accordance with prespecification" (2020, 477). An example of bad practice would be preregistering a trial with a particular design while reporting the results that are generated by using a different design. Challenges connected to reporting primarily concern the preference for reporting only positive results, thereby withdrawing important information about negative results from clinicians and health systems. Another challenge is ascertaining quality control because expert reviewers are a scarce resource. The last condition, feasibility, is especially challenging in a crisis. London and Kimmelman argue that this nonetheless should not mean that it is justifiable to trade it off against the other four conditions. An increase in feasibility does not mean a decrease in addressing important evidence gaps, allowing less rigorous design, neglecting analytical integrity, or failing to transparently report. They give particular guidelines to clinicians:

> Individual clinicians should avoid off-label use of unvalidated interventions that might interfere with trial recruitment and resist the urge to carry out uncontrolled, open-label studies. They should instead seek out opportunities to join larger, carefully orchestrated protocols to increase the prospect that high-quality studies will be completed quickly and generate the information needed to advance individual and public health. Academic medical centers can facilitate such coordination by surveying the landscape of ongoing studies and establishing mechanisms for "prioritization review" to triage studies. (London and Kimmelman 2020, 477)

Channeling resources to orchestrated endeavors is a result of decades-long efforts to transform biomedical research towards better coordination and private-public partnerships, against the backdrop of the big data era that brought along the need to store, manage, and adequately use vast amounts of information and material. This portrays a picture in which the key to upholding standards for implementing adaptive design trials is in the hands of big research organizations with enough infrastructure and resources to

embark on such a complex task. I call this *operational exceptionalism*, in which centralization and coordination are the prerequisites for simultaneously increasing both the speed of generating evidence and the quality of this evidence. The only way to counter pandemic research exceptionalism seems to be by endorsing operational exceptionalism, according to which adaptive trials are not useful when run autonomously in local settings, but only when they are a part of larger projects based in selected research institutions.

4. Adaptive Trials and Precision Medicine

In this section, I focus on adaptive design as a clinical trial innovation that followed up on novel research methods and increased understanding of the intervention that is being assessed. In this cluster of cases, adaptive design trials are related to the rise of precision medicine.

Personalized or precision medicine[2] is an approach that tailors therapy to individual needs. It is often represented as 'P4' medicine: predictive, preventive, personalized, and participatory. The observations of highly variable drug responses have led to the development of a new scientific discipline from genetics, biochemistry, and pharmacology, namely pharmacogenetics, while advances in molecular medicine have led to a pharmacogenomics which seeks to understand the molecular mechanisms of drug response (Vogenberg, Barash, and Pursel 2010). In this new approach, patients' gene variations guide the selection and dosage of drugs. Several adaptive measures have been introduced to evaluate precision medicine treatments and to match the well-responding subgroups of patients with promising therapies, improve access, and evaluate efficacy earlier and more efficiently.

An example of an adaptive trial for a precision medicine intervention is the BATTLE-2 study—The Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination 2 (Garralda et al. 2019). Results generated in the 'adaptive phase' inform the randomization to different drugs or combinations based on mutation profiles.

[2] Terms 'personalized' and 'precision' medicine are often used interchangeably, although personalized medicine is the older term, while precision medicine is currently the preferred one, at least according to the US National Research Council (NRC). NRC adopts the following definition of both terms: "the tailoring of medical treatment to the individual characteristics of each patient (…) to classify to a specific treatment" (NRC 2011, 12). 'Precision medicine' is preferred to avoid the interpretation that 'personalized' means that each patient will be treated differently.

> Instead of using a fixed model—built on the training data only–adaptive strategies use the information on patients enrolled earlier in the testing set to continuously update the model and refine accrual throughout the entire study. (Garralda et al. 2019, 551)

Accrual design is a type of adaptive design—after the initial 'learning phase', in the 'adaptive phase' the ratio of patients randomly assigned to the experimental arm as opposed to the control arm changes to increase the proportion of patients in the arm that is performing better, which also increases the statistical power to detect clinical benefit (Garralda et al. 2019, 551). Adaptive enrichment is a term that refers to the modification of the patient eligibility criteria: if analysis shows that one subgroup has a more favorable response, the trial can be 'enriched' by modifying it to either exclusively or predominantly enroll patients from this subgroup (Thorlund et al. 2018). The seamless adaptive trial design allows for proceeding from phase II to phase III trial in a non-standard way. The results from the phase II trial are used to determine the initial patient allocation ratio, the planned total sample size (which can be rather smaller than the usual phase III samples that normally include from 300 to several thousand patients), and a potentially enriched set of patients, those that are thought to benefit the most from the intervention (Thorlund et al. 2018).

A significant part of the literature on adaptive trials, including guidelines for their implementation and reporting, comes from precision medicine research groups. They are raising problems related to their usage, but also providing means of addressing and overcoming them (for example, Garralda et al. 2019; Pallmann et al. 2018). Each trial is adapted in a particular way, so informed consent and the effective communication of risks and benefits to the patients can be a problem (Garralda et al. 2019, 552). Funders are suspicious about the validity of adaptive trials or lack experience in evaluating them, so may decide against approving them (Garralda et al. 2019; Pallmann et al. 2018). Regulators alike may be unfamiliar with adaptive design (Pallmann et al. 2018, 4). Operational challenges such as managing preplanned adaptations together with blinding may require additional staff and experience, as data may leak more easily and reach the sponsors, compromising the integrity of the trial (Pallmann et al. 2018, 5).

Overall, the efficacy of adaptive trials can be uncertain due to many factors, which are often local, contingent, and practical. Advocates of the usage of adaptive trials argue that these problems can be countered by transparent planning, careful execution, and the rigorous interpretation of the results. Additional skills in planning, conducting, and analyzing

adaptive design trials would need to be at hand, including statistical, mathematical, and modeling expertise. Since many clinicians are not trained in their usage, while the regulators are uncertain about their potential to avoid problems that the standard randomization and bias-reducing measures are in place for, their wider usage is both called for and cautioned against, sometimes by the very same authors (like Pallmann et al. 2018 from the clinical medicine side) and regulatory documents (FDA 2019). On the cautious side, it is emphasized that randomization and blinding remain the most reliable indicators of objectivity in clinical research and should not be bypassed in favor of shorter trials. A particularly problematic practice is reliance on non-randomized and non-blinded studies, and avoidance of control groups. On the affirmative side, novel designs such as multi-arm and seamless design trials are characterized as being a well-understood, ethical and efficient way of doing clinical research.

5. Adaptive Trials and the Productivity Crisis

From another vantage point, the pharmaceutical industry is voicing hopes about the usage of adaptive trials as a means to end the productivity crisis (Mahlich, Bartol, and Dheban 2021). In this section, I place the emergence of adaptive trials in a wider context of biomedical movements initiated to improve the productivity and cost-benefit of biomedical research.

Existing resources for the implementation of adaptive trials are a product of diverse measures in place to reform the pace and path by which biomedical innovations reach the market and patients. There is a consensus that pharmaceutical productivity has been going through a crisis for at least three decades (Munos 2009; Pammolli, Magazzini, and Riccaboni 2011; Taylor 2016). Advances in basic science resulting from stem cell research and the Human Genome Project (completed in 2003) have not resulted in clinical applications as quickly as was initially expected (Solomon 2015, 161-163). The so-called 'pipeline problem' refers to the slowdown, instead of the expected acceleration, in innovative medical therapies reaching patients (FDA 2004), and what has thus been sought is the 'uncorking of the bottleneck' of pharmaceutical innovation. Furthermore, it has been estimated that it takes 17 years on average for research results to find implementation in clinical practice, which has been considered too slow (Morris et al. 2011). These problems have motivated different initiatives to transform the way biomedical research is conducted. Consequently, in the 2000s the idea of 'translational research' became a "buzzword" (Fishburn 2013, 487), a "mantra" (Maienschein et al. 2008, 43), "in vogue"

(Fang and Casadevall 2010, 563), and even "an imperative" (Harrington and Hauskeller 2014).

The translational approach is based on the prospect of directly matching ideas for new therapies with the needs of patients observed in the clinic. It can be described as a cluster of accelerated transitions in the development of a medical product at the intersection of basic and clinical research, and more broadly, the intersection of prevention, guidelines, and health policy. These transitions are mostly accelerated by external, non-scientific measures: better communication between researchers from different disciplines, better communication between different stakeholders such as patients, researchers, regulators, and producers of therapies, interdisciplinary training, collection of databanks, and building of new research centers that would facilitate the interaction between basic and clinical research. Most of the philosophical work on translational medicine shares the view that it is hard to "find substance amidst the rhetoric" and that the movement "appears to offer no more than a metaphor" (Fuller 2016).

Robinson (2019) pointedly argues that attempts to find epistemic novelty in the new medical movements fail because their objectives are better assessed by a social epistemology approach attentive to market forces and financialized models of science and innovation.

> TrM (translational medicine) cannot be analyzed merely in terms of its epistemic novelty. After all, it has relocated research practices from the R&D departments of biopharmaceutical partners to university laboratories. (…) It is—in its current functionality—a structural configuration for the externalization of the costs and risks of early-stage biopharmaceutical research and development onto universities. (Robinson 2019, 4404)

Translational initiatives are thus comprised of "questions, methods, areas of concern, and projects" which are "a product of a specific set of financial, commercial and industry-driven shifts" (Robinson 2019, 4404).

Justification in terms of patient empowerment and acceleration of discovery and research is shared in both translational and precision initiatives. Both movements value speed in discovery, research, and development, which is not only a success of science but of a larger cooperative work and exchange of many stakeholders, institutions, and disciplinary cultures. Finally, it was the biobanks collected as part of translational initiatives in the early 2000s that have made it possible to

personalize medicine in the 2010s.[3] Contemporary translations are very likely to occur on the terrain of precision medicine and they occur there faster due to changes in drug discovery methods and clinical assessment routes.[4] In drug discovery, methods such as high-throughput screening can identify molecular targets among a vast number of potential matches (Adam 2011), and in clinical assessment, the adaptive design facilitates matching subgroups of patients with promising therapies based on genetic profiling.

Against this backdrop, the emergence and development of adaptive designs can be traced to translational and precision medicine centers. Increased awareness of the need for trained statisticians, mathematicians, and big-data experts in clinical research teams, and opening up to interdisciplinarity in a variety of contexts where singular expertise is not sufficient, have contributed to the fact that adaptive trials are nowadays planned, conducted, analyzed, and regulated with more understanding and expertise. However, this fact alone does not grant justification for their usage in every instance of clinical research. Clear rationale, transparent protocols, and importantly, operational conditions, need to be in place. It seems that especially operational conditions cannot be satisfied on smaller scales of individual clinics and local research centers, but rather "orchestrated" by big consortia with sufficient resources and in close cooperation with policy makers and industrial partners. The complexities that this operational exceptionalism brings in a value-laden and interest-driven environment of biomedical research are beyond the scope of this paper but call for attention and discussion by philosophers and social scientists alike.

6. Conclusion

The success of Covid-19 adaptive trials is not a consequence of research exceptionalism or lucky guesses, but of prior experience in healthcare crisis-management and structured efforts to reform biomedical research and innovation. That said, it is important to qualify the context in which adaptive trials are conducted and implemented. It is a private-public partnership of many stakeholders, highly burdened with both social commitments and commercial interests. Importantly, the apparent flexibility of adaptive trials is not as flexible as it may seem at first sight.

[3] Initiatives such as the NIH Roadmap in the US (NIH 2014) and the reforms outlined in the Cooksey Report (2006) in the UK.

[4] In 2017 the number of FDA approvals hit a two-decade high with 46 novel medicines, followed by 59 approvals in 2018 (Mullard 2019). More precision medicines and tests were approved in 2017 than any year before (Bilkey et al. 2019), many of them based on biomarkers reliant on genetic testing.

They require both planning and rigor to be successful, just as much as fixed trials. The usual standards of rigor remain unchallenged in the new context, coming down to blinding, randomization, and controls. A new and most valuable element of their success is their speed. However, it is a qualified speed that, rather than trading off against reliability, requires reliability to achieve epistemic benefit. Daniel Steel (2010, 26-28) would call it an extrinsically epistemic value, i.e. a value that is not truth conducive *per se* but in combination with an intrinsically epistemic value like accuracy. Adaptive designs ground their reliability in "orchestration" and integration of different evidence and expertise. In the case of clinical trials, the benefits are both ethical—earlier access to therapies, and epistemological—earlier results that inform policies and further research. Still, adaptive design trials require additional resources and coordination, which is the most pressing practical obstacle to their wider, local implementation. They have been increasingly developed as a part of the precision medicine approach, and have recently been used to assess Covid-19 therapies. It is important to keep in mind though, that this does not grant them the status of the new standard. It means at best that the standard welcomes necessary upgrades and contextual adjustments.

Acknowledgments

I would like to thank two anonymous reviewers for their extremely helpful comments and suggestions, and Lucie White for her detailed feedback on the revised version of the paper.

REFERENCES

Berry, Donald A. 2011. "Adaptive Clinical Trials: The Promise and the Caution." *Journal of Clinical Oncology* 29 (6): 606–9. https://doi.org/10.1200/JCO.2010.32.2685.

Biddle, Justin. 2007. "Lessons from the Vioxx Debacle: What the Privatization of Science Can Teach Us about Social Epistemology." *Social Epistemology* 21 (1): 21–39. https://doi.org/10.1080/02691720601125472.

Bilkey, Gemma A., Belinda L. Burns, Emily P. Coles, Trinity Mahede, Gareth Baynam, and Kristen J. Nowak. 2019. "Optimizing Precision Medicine for Public Health." *Frontiers in Public Health* 7: 42. https://doi.org/10.3389/fpubh.2019.00042.

Branch-Elliman, Westyn, A Rani Elwy, and Paul Monach. 2020. "Bringing New Meaning to the Term 'Adaptive Trial': Challenges of Conducting Clinical Research during the Coronavirus Disease

2019 Pandemic and Implications for Implementation Science." *Open Forum Infectious Diseases* 7 (11). https://doi.org/10.1093/ofid/ofaa490.

Calain, Philippe. 2018. "The Ebola Clinical Trials: A Precedent for Research Ethics in Disasters." *Journal of Medical Ethics* 44 (1): 3–8. https://doi.org/10.1136/medethics-2016-103474.

Cooksey, Sir David. 2006. *A review of UK health research funding*. London: Crown, HM Treasury. Accessed November 12, 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/228984/0118404881.pdf

Crewe, Tom. 2018. 'Here was a plague. Review of the books *How to Survive a Plague: The Story of How Activists and Scientists Tamed Aids*, *Patient Zero and the Making of the Aids Epidemic*, *Modern Nature: The Journals of Derek Jarman, 1989–90*, *Smiling in Slow Motion: The Journals of Derek Jarman, 1991–94* and *The Ward*'. *London Review of Books* 40 (18): 7–16. Accessed September 3, 2021. https://www.lrb.co.uk/the-paper/v40/n18/tom-crewe/here-was-a-plague

Epstein, Steven. 1996. *Impure Science: AIDS, Activism, and the Politics of Knowledge*. Reprint. Medicine and Society 7. Berkeley, Calif.: University of California Press.

European Medicines Agency (EMA). 2017. *Adaptive pathways*. Accessed November 12, 2021. https://www.ema.europa.eu/en/human-regulatory/research-development/adaptive-pathways

Fang, Ferric C., and Arturo Casadevall. 2010. "Lost in Translation—Basic Science in the Era of Translational Research." *Infection and Immunity* 78 (2): 563–66. https://doi.org/10.1128/IAI.01318-09.

Fishburn, C. Simone. 2013. "Translational Research: The Changing Landscape of Drug Discovery." *Drug Discovery Today* 18 (9–10): 487–94. https://doi.org/10.1016/j.drudis.2012.12.002.

Food and Drug Agency (FDA). 2004. *Innovation/Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products*. Accessed September 3, 2021. https://www.who.int/intellectualproperty/documents/en/FDAproposals.pdf

Food and Drug Agency (FDA). 2019. *Adaptive Design Clinical Trials for Drugs and Biologics. Guidance for Industry*. Accessed November 12, 2021. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry

Fuller, Jonathan. 2016. "The Reading Room: 'Making Medical Knowledge'". *British Medical Journal* blogs. Accessed November 12, 2021.

https://blogs.bmj.com/medical-humanities/2016/04/25/the-reading-room-making-medical-knowledge/

Garralda, Elena, Rodrigo Dienstmann, Alejandro Piris-Giménez, Irene Braña, Jordi Rodon, and Josep Tabernero. 2019. "New Clinical Trial Designs in the Era of Precision Medicine." *Molecular Oncology* 13 (3): 549–57. https://doi.org/10.1002/1878-0261.12465.

Geisbert, Thomas W. 2017. "First Ebola Virus Vaccine to Protect Human Beings?" *The Lancet* 389 (10068): 479–80. https://doi.org/10.1016/S0140-6736(16)32618-6.

Harrington, Jean, and Christine Hauskeller. 2014. "Translational Research: An Imperative Shaping the Spaces in Biomedicine." *Tecnoscienza: Italian Journal of Science & Technology Studies* 5 (1): 191–202.

Henao-Restrepo, Ana Maria, Anton Camacho, Ira M. Longini, Conall H. Watson, W. John Edmunds, Matthias Egger, Miles W. Carroll, et al. 2017. "Efficacy and Effectiveness of an RVSV-Vectored Vaccine in Preventing Ebola Virus Disease: Final Results from the Guinea Ring Vaccination, Open-Label, Cluster-Randomised Trial (Ebola Ça Suffit!)." *The Lancet* 389 (10068): 505–18. https://doi.org/10.1016/S0140-6736(16)32621-6.

London, Alex John, and Jonathan Kimmelman. 2020. "Against Pandemic Research Exceptionalism." *Science* 368 (6490): 476–77. https://doi.org/10.1126/science.abc1731.

Longini, Ira M., John-Arne Røttingen, Marie Paule Kieny, W. John Edmunds, and Ana Maria Henao-Restrepo. 2018. "Questionable Efficacy of the RVSV-ZEBOV Ebola Vaccine – Authors' Reply." *The Lancet* 391 (10125): 1021–22. https://doi.org/10.1016/S0140-6736(18)30559-2.

Mahlich, Jörg, Arne Bartol, and Srirangan Dheban. 2021. "Can Adaptive Clinical Trials Help to Solve the Productivity Crisis of the Pharmaceutical Industry? - A Scenario Analysis." *Health Economics Review* 11 (1): 4. https://doi.org/10.1186/s13561-021-00302-6.

Maienschein, Jane, Mary Sunderland, Rachel A. Ankeny, and Jason Scott Robert. 2008. "The Ethos and Ethics of Translational Research." *The American Journal of Bioethics* 8 (3): 43–51. https://doi.org/10.1080/15265160802109314.

Matthias, Adam. 2011. "Multi-Level Complexities in Technological Development: Competing Strategies for Drug Discovery." In *Science in the Context of Application*, edited by Martin Carrier and Alfred Nordmann, 274:67–83. Boston Studies in the Philosophy of Science. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-9051-5.

Metzger, Wolfram G., and Sarai Vivas-Martínez. 2018. "Questionable Efficacy of the RVSV-ZEBOV Ebola Vaccine." *The Lancet* 391 (10125): 1021. https://doi.org/10.1016/S0140-6736(18)30560-9.

Morris, Zoë Slote, Steven Wooding, and Jonathan Grant. 2011. "The Answer Is 17 Years, What Is the Question: Understanding Time Lags in Translational Research." *Journal of the Royal Society of Medicine* 104 (12): 510–20. https://doi.org/10.1258/jrsm.2011.110180.

Mullard, Asher. 2019. "2018 FDA Drug Approvals." *Nature Reviews Drug Discovery* 18 (2): 85–89. https://doi.org/10.1038/d41573-019-00014-x.

Munos, Bernard. 2009. "Lessons from 60 Years of Pharmaceutical Innovation." *Nature Reviews Drug Discovery* 8 (12): 959–68. https://doi.org/10.1038/nrd2961.

National Institute of Health (NIH). 2014. *A Decade of Discovery: The NIH Roadmap and Common Fund (2004-2014)*. Accessed November 12, 2021. https://commonfund.nih.gov/sites/default/files/ADecadeofDiscoveryNIHRoadmapCF.pdf

National Research Council (U.S.). 2011. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*. Washington, D.C: National Academies Press.

Pallmann, Philip, Alun W. Bedding, Babak Choodari-Oskooei, Munyaradzi Dimairo, Laura Flight, Lisa V. Hampson, Jane Holmes, et al. 2018. "Adaptive Designs in Clinical Trials: Why Use Them, and How to Run and Report Them." *BMC Medicine* 16 (1): 29. https://doi.org/10.1186/s12916-018-1017-7.

Pammolli, Fabio, Laura Magazzini, and Massimo Riccaboni. 2011. "The Productivity Crisis in Pharmaceutical R&D." *Nature Reviews Drug Discovery* 10 (6): 428–38. https://doi.org/10.1038/nrd3405.

Riley, William T. 2016. "Chapter 18 - a New Era of Clinical Research Methods in a Data-Rich Environment." In *Oncology Informatics*, edited by Bradford W. Hesse, David K. Ahern, and Ellen Beckjord, 343–55. Boston: Academic Press. https://doi.org/10.1016/B978-0-12-802115-6.00018-5.

Robinson, Mark D. 2019. "Financializing Epistemic Norms in Contemporary Biomedical Innovation." *Synthese* 196 (11): 4391–4407. https://doi.org/10.1007/s11229-018-1704-0.

Solomon, Miriam. 2015. *Making Medical Knowledge*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198732617.001.0001.

Sonnabend, Jospeh. 2011. 'Remembering the Original AZT Trial'. POZ Magazine Online. Accessed November 12 2021. https://www.poz.com/blog/-v-behaviorurldefaul

Stallard, Nigel, Lisa Hampson, Norbert Benda, Werner Brannath, Thomas Burnett, Tim Friede, Peter K. Kimani, et al. 2020. "Efficient Adaptive Designs for Clinical Trials of Interventions for Covid-19." *Statistics in Biopharmaceutical Research* 12 (4): 483–97. https://doi.org/10.1080/19466315.2020.1790415.

Steel, Daniel. 2010. "Epistemic Values and the Argument from Inductive Risk." *Philosophy of Science* 77 (1): 14–34. https://doi.org/10.1086/650206.

Taylor, David. 2015. "The Pharmaceutical Industry and the Future of Drug Development." In *Pharmaceuticals in the Environment*, edited by Ronald E. Hester and Roy M. Harrison, 1–33. The Royal Society of Chemistry. https://doi.org/10.1039/9781782622345-00001.

Thorlund, Kristian, Jonas Haggstrom, Jay JH Park, and Edward J. Mills. 2018. "Key Design Considerations for Adaptive Clinical Trials: A Primer for Clinicians." *BMJ* 360: k698. https://doi.org/10.1136/bmj.k698.

Upshur, Ross, and Jonathan Fuller. 2016. "Randomized Controlled Trials in the West African Ebola Virus Outbreak." *Clinical Trials* 13 (1): 10–12. https://doi.org/10.1177/1740774515617754.

Varghese, Joby. 2021a. "Influence and Prioritization of Non-Epistemic Values in Clinical Trial Designs: A Study of Ebola Ça Suffit Trial." *Synthese* 198 (10): 2393–2409. https://doi.org/10.1007/s11229-018-01912-0.

———. 2021b. "Non-Epistemic Values in Shaping the Parameters for Evaluating the Effectiveness of Candidate Vaccines: The Case of an Ebola Vaccine Trial." *History and Philosophy of the Life Sciences* 43 (2): 1–15. https://doi.org/10.1007/s40656-021-00417-3.

Vogenberg, F. Randy, Carol Isaacson Barash, and Michael Pursel. 2010. "Personalized Medicine." *Pharmacy and Therapeutics* 35 (10): 560–76.

Wilholt, Torsten. 2009. "Bias and Values in Scientific Research." *Studies in History and Philosophy of Science Part A* 40 (1): 92–101. https://doi.org/10.1016/j.shpsa.2008.12.005.

World Health Organization (WHO). 2015. 'Ring Vaccination Trial Consortium Ebola ça Suffit! Ebola vaccine Phase 3 trial Guinea'. Accessed September 3, 2021. http://www.who.int/immunization/research/meetings_workshops/3_Ebola_Ring_vaccination_Phase_3_trial_2015.pdf

European Journal of Analytic Philosophy EuJAP | Vol. 17 | No. 2 | 2021
https://doi.org/10.31820/ejap.17.3.3

WRONGFUL MEDICALIZATION AND EPISTEMIC INJUSTICE IN PSYCHIATRY: THE CASE OF PREMENSTRUAL DYSPHORIC DISORDER

Anne-Marie Gagné-Julien[1]

[1] Biomedical Ethics Unit, McGill University

Original scientific article – Received: 30/4/2021 Accepted: 13/10/2021

ABSTRACT

In this paper, my goal is to use an epistemic injustice framework to extend an existing normative analysis of over-medicalization to psychiatry and thus draw attention to overlooked injustices. Kaczmarek (2019) has developed a promising bioethical and pragmatic approach to over-medicalization, which consists of four guiding questions covering issues related to the harms and benefits of medicalization. In a nutshell, if we answer "yes" to all proposed questions, then it is a case of over-medicalization. Building on an epistemic injustice framework, I will argue that Kaczmarek's proposal lacks guidance concerning the procedures through which we are to answer the four questions, and I will import the conceptual resources of epistemic injustice to guide our thinking on these issues. This will lead me to defend more inclusive decision-making procedures regarding medicalization in the DSM. Kaczmarek's account complemented with an epistemic injustice framework can help us achieve better forms of medicalization. I will then use a contested case of medicalization, the creation of Premenstrual Dysphoric Disorder (PMDD) in the DSM-5 to illustrate how the epistemic injustice framework can help to shed light on these issues and to show its relevance to distinguish good and bad forms of medicalization.

Keywords: *over-medicalization; epistemic injustice; premenstrual dysphoric disorder; hermeneutical injustice; pre-emptive testimonial injustice; Miranda Fricker*

Correspondence: annemariegagnejulien@gmail.com

Introduction

Medicalization is a controversial topic both within and outside psychiatry, especially since the publication of the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5, APA 2013). Several critics have argued that the DSM-5 medicalizes conditions that should only be considered normal life problems (e.g., Lane 2007; Frances 2010, 2013; Horwitz and Wakefield 2012; see also Stegenga 2021 and Murphy-Hollies 2021 in this issue of EuJAP). However, although medicalization in psychiatry is generally discussed from a critical perspective, the term itself is value-neutral: from a sociological point of view, medicalization can bring both good and bad consequences (e.g., Conrad et al. 2010). What appears problematic are the bad forms of medicalization, or what has been called "over-medicalization" (e.g., Conrad 2013; Conrad and Slodden 2013). Regarding the many consequences and implications of medicalization, identifying cases of medicalization from an ethical point of view is a difficult undertaking. Some philosophers and ethicists have recently taken up this ambitious task (e.g., Parens 2013; Kaczmarek 2019), but have not reached a consensus.

In parallel, the framework of epistemic injustices (hereafter EI) as developed by Miranda Fricker (2007, 2017) has proven fruitful in psychiatry and mental health care. EI are the harms suffered by individuals belonging to oppressed groups in their capacities as epistemic agents, due to prejudicial identity stereotypes or to the marginalization associated with these groups. These injustices can arise at various points in the process of knowledge acquisition and transmission, such as interpreting an experience or offering a testimony.

Where medicine is concerned, Kidd and Carel (2017; see also Carel and Kidd 2014, 2016, 2018, forthcoming) have depicted a particular form of EI that concern prejudices associated with the experience of illness, called *pathocentric epistemic injustices*. Pathocentric epistemic injustices occur when

> ill persons [are] being ignored, silenced, or dismissed; [are] not being listened to or taken seriously, and [are] being treated as mere sources of information, only able to answer within the defined terms of clinical-epistemic practice. (Kidd and Carel forthcoming)

As some have argued, the risk of encountering this type of EI is even greater in psychiatry because of widespread negative stereotypes associated with mental illness (Crichton et al. 2017; see also e.g., Kurs and

Grinshpoon 2018; Kyratsous and Sanati 2015; LeBlanc and Kinsella 2016). The application of the conceptual framework of EI has thus made it possible to target various ethical problems related to knowledge production and transmission in psychiatry (e.g., Kyratsous and Sanati 2017; Crichton et al. 2017; Kurs and Grinshpoon 2017; Tate 2018; Gosselin 2018; Bueter 2019; Sullivan 2019).

In this paper, my goal is to use the EI framework to extend an existing normative analysis of over-medicalization to psychiatry and thus draw attention to overlooked injustices. Kaczmarek (2019) has developed a promising bioethical and pragmatic approach to over-medicalization, which consists of four guiding questions covering issues related to the harms and benefits of medicalization. In a nutshell, if we answer "yes" to all proposed questions, then it is a case of over-medicalization. Building on the EI framework, I will argue that Kaczmarek's proposal lacks guidance concerning the procedures through which we are to answer the four questions, and I will import the conceptual resources of EI to guide our thinking on these issues. This will lead me to defend more inclusive decision-making procedures regarding medicalization in the DSM. Kaczmarek's account complemented with the EI framework can help us achieve better forms of medicalization.

The paper is organized as follows: in section 1, I will first define medicalization and introduce the challenge of "wrongful medicalization", i.e. the task of distinguishing good and bad forms of medicalization. Secondly, I will critically review previous accounts which have tried to overcome this challenge. I will argue that Kaczkmarek's proposal is a promising one, but needs to be further developed. In section 2, I will suggest that the EI framework draws attention to some overlooked ethical wrongs related to medicalization, if we understand the medicalization process as a transformation of hermeneutical resources implying power relations between different actors. I will then argue that the EI framework should complement Kaczmarek's account in order to reduce the risk of epistemic injustices induced by medicalization, and therefore the risk of wrongful medicalization. In section 3, to illustrate the relevance of my proposal, I will apply this conclusion to a case study: the medicalization of Premenstrual Dysphoric Disorder (PMDD) in DSM-5. I will then suggest possible improvements based on the findings of Section 2.

1. Medicalization in Psychiatry and the Bioethical Challenge of "Wrongful Medicalization"

1.1 The Social Process of Medicalization in Psychiatry: Some Methodological Notes

"Medicalization"[1] does not always have the same meaning in the literature (for review, see e.g., Davies 2010; Hofmann 2016; Busfield 2017). In this paper, I will use the following broad definition:

> Medicalization occurs when previously nonmedical problems become defined (and treated) as medical problems, usually as an illness or disorder. (Conrad and Slodden 2013, 62)

While this broad definition can encompass a large array of phenomena, I will restrict my analysis to a specific context, i.e. that of North American contemporary psychiatry. In this context, medicalization generally occurs through the revision of the official nosological manual, the *Diagnostic and Statistical Manual* (DSM) published by the American Psychiatric Association (APA). Moreover, in what follows, I will focus on two main actors of the medicalization process: people living with mental illness and the main North American psychiatric institutions by which medicalization occurs, the APA (and the revision structures of the DSM). It is important to recognize that there are other actors involved in this process (e.g., pharmaceutical industries, other healthcare professionals, laypeople, the media, etc.) and other contexts in which medicalization happens (the globalization of medical concepts, the rest of medicine, etc.), but the scope of this paper does not allow me to cover them all in detail.

One way for medicalization to happen in North American psychiatry is through the categorization of a condition as a new mental disorder in the DSM. A paradigmatic example is the creation of Post-Traumatic Stress Disorder (PTSD) in the DSM-III (APA 1980). Despite controversies about its existence as a distinct diagnosis, PTSD was introduced in the DSM following pressure from anti-war psychiatrists and Vietnam veterans who were experiencing symptoms of trauma, such as flashbacks and intense anxiety (see e.g., Scott 1990; Riska 2013). Another, more recent example, on which I will return in section 3 of the paper, is the medicalization of Premenstrual Dysphoric Disorder (PMDD), a new diagnostic category

[1] Although the trend in psychiatry is toward increased medicalization, a condition can, conversely, be removed from the medical field. This phenomenon is called "demedicalization". For example, homosexuality was excluded from the DSM and thus from the medical field following demands by groups campaigning for homosexual rights (APA, 1973, for a detailed discussion, see e.g., Kirk and Kutchins 1992).

introduced in the DSM-5 (2013). PMDD refers to the distress associated with the menstrual cycle in menstruating women and is considered to be a more extreme form of Premenstrual Syndrome (PMS). Some feminist critics welcomed the new diagnosis with contention, worrying, among other things, about the illegitimate pathologization of women's anger.

Although medicalization generally refers to such a process, i.e. in which a non-medical condition is transformed into a medical category, it can also occur through the revision of already-existing diagnoses. Taken in this latter sense, medicalization happens when individuals who were not diagnosed with a mental disorder become so when the clinical description of the diagnostic criteria changes. That is, when specific diagnostic criteria are modified, when criteria thresholds are revised, or when new age ranges are included in them. Such cases do not involve the creation of new psychiatric categories, but only the expansion of already-existing ones (Conrad and Slodden 2013, 65). A good example of a controversial case of this type of medicalization is Major Depressive Disorder (MDD), and more specifically the debate surrounding the removal of the bereavement exclusion criterion in the DSM-5.[2] In the DSM-IV, people suffering from depressed mood caused by the loss of a loved one were not diagnosed with MDD if the sadness experienced was proportionate to the loss. In the DSM-5, the bereavement clause was removed (APA 2013, 161). A person can now be diagnosed with MDD if she meets MDD diagnostic criteria, despite grief being the cause of her symptoms. According to some critics, this could lead to an increase in the prevalence of the disorder. Worse: it could mean diagnosing people with a mental disorder while they suffer from normal sadness associated with the grieving process (for a more detailed discussion, see e.g., Horwitz and Wakefield 2007; Pies 2014; Bandini 2015).

1.2 The Problem of Wrongful Medicalization in Psychiatry

Historically, the term "medicalization" is connected with the work of famous critics of psychiatry and medicine such as Thomas Szasz, Ivan Illich, and Irving Zola, who pointed out the illegitimate hold or social control exerted by medical institutions over "deviance" (or what was perceived as such). However, contemporary critics have recently started to restrict the scope of their criticism to specific diagnoses, arguing that only

[2] Other instances of this type of medicalization include the diagnosis of Bipolar Disorder in children (BD, see e.g., Healy 2008) or the diagnosis of Attention Deficit Hyperactivity Disorder (ADHD) in adults (e.g., Conrad 2007; Conrad and Slodden 2013). In both cases new individuals are medicalized because of a change in age ranges and age-related diagnostic criteria. Another way in which medicalization can happen is via the general definition of mental disorder in the DSM (see e.g., Cooper 2015).

these would be illegitimate forms of medicalization (e.g., Charland 2013; Sedler 2015). Moreover, despite the numerous criticisms aimed at medicalization, most sociologists take the process to be value-neutral. Medicalization is understood as a social process that can bring *both* positive and negative consequences for individuals and society (Conrad 2007). The benefits of medicalization include granting better access to care, motivating people to look for help and resources, decreasing blame associated with medicalized conditions, etc. Disadvantages include depreciating the importance of social context in explications of mental distress, medicalizing all domains of human life to create a unilateral, purely medical understanding of normality, spawning unnecessary clinical interventions, generating high costs in public health care systems, etc.[3] Medicalization is thus neither an inherently negative nor an inherently positive process, making the ethical assessment of it difficult.

Therefore, the literature generally does not discuss medicalization itself, but rather what has been called "over-medicalization" (Conrad and Slodden 2013; Conrad 2013). Over-medicalization usually refers to the process of "altering the meaning or understanding of experiences, so that human problems are reinterpreted as medical problems requiring medical treatment, *without net benefit to patients or citizens*" (Carter et al. 2015, table 1, emphasis added). In other words, "over-medicalization" is often used when conditions are believed to have been unnecessarily, wrongfully, or even harmfully medicalized.[4] However, since medicalization brings both positive and negative consequences, drawing the line between the good and bad forms of this social process is extremely complex. Psychiatry is often faced with practical problems, like whether particular diagnoses should be included in the DSM (e.g., should PTSD or PMDD be included in the DSM?), or whether specific diagnostic criteria for existing diagnoses should be modified (e.g., should the bereavement exclusion criteria be kept or removed from the clinical description of MDD?). The issue here is rather to distinguish cases in which psychiatry expands its domain within its legitimate scope, and other cases in which such expansion proves excessive (see e.g., Purdy 2001; Sadler et al. 2009; Reiheld 2010; Parens

[3] For a more detailed discussion of the advantages and disadvantages of medicalization, see e.g., Stein et al. (2006), Davis (2010), Reiheld (2010), Bastra and Frances (2012), Parens (2013), Conrad and Slodden (2013), Kaczmarek (2019), and Thomas (2021).

[4] "Overdiagnosis" is also used about cases in which an existing diagnosis is applied to a condition with few or no symptoms (e.g., Moynihan et al. 2012; but see Rogers and Mintzker 2016 for distinctions). "Disease mongering" is sometimes used as well to describe situations in which the pharmaceutical industry influences the expansion of the medical field (e.g., Moynihan et al. 2002; Moynihan and Cassels 2005). Overdiagnosis and disease mongering are thus specific manifestations of over medicalization, the latter referring to the more general phenomenon by which the medical field expands (for the opposite view, see Hoffman 2016).

2013; Murano 2018; Kaczmarek 2019; see also Carter et al. 2015, 2016 on overdiagnosis specifically).

One strategy to assess whether a case results from over-medicalization involves arguing that a condition has been wrongfully introduced in medical classification. That is, the condition is not "truly" medical, and has been mistakenly understood as such. I will call this approach the "substantive account". In philosophy of psychiatry, the work of Horwitz and Wakefield (2007; see also e.g., Boorse 1976 for a similar point), among others, belongs to this approach. Horwitz and Wakefield's strategy is to appeal to a scientific or objective component to draw the line between good and bad forms of medicalization. According to their account, mental disorders are harmful dysfunctions.[5] They argue that psychiatry should restrict the scope of the concept of mental disorder to harm-inducing deviations from the evolving norms of mental functioning. Within this framework, over-medicalization happens when psychiatry does not refer to the natural and objective definition of mental disorder and extends beyond the scope of this definition. Horwitz and Wakefield focus primarily on the diagnosis of MDD, arguing that the DSM is overly inclusive about some forms of normal sadness. This excess results in the false diagnosis of healthy individuals.

While promising, Horwitz and Wakefield's strategy is not without problems. Very briefly, their approach is limited by the speculative nature of an evolutionary definition of mental dysfunction and by its vague notion of harm. Although the evolution of the human mind is not what is at stake here, the state of our knowledge about the traits and mechanisms selected for in past human history is too poor to allow us to distinguish mental disorders from normal mental functioning in practical situations (e.g., Lilienfeld and Marino 1995; Murphy and Woolfolk 2000; McNally 2001; Schramme 2010; Bingham and Banner 2014; Faucher 2021). Moreover, although "harm" seems like a good fit here, the notion is underspecified in Wakefield's definition, since it is not clear how we are supposed to apply this criterion in real-life situations (e.g., Powell and Scarffe 2019 a,b; De Block and Sholl 2021; see, however, Wakefield and Conrad 2019 for a response). In its current state, Horwitz and Wakefield's account is very difficult to use if we want to identify cases of over-medicalization.

In contrast to "substantive" accounts of over-medicalization—and because of their limitations—many authors have argued that the definition of what constitutes a mental disorder and the establishment of proper boundaries

[5] Note that this account has been initially developed by Wakefield, see e.g., Wakefield (1992, 1999).

for psychiatry are fundamentally normative issues (e.g., Cooper 2005, Conrad and Barker 2010).[6] The medicalization of health conditions appears as a value-laden process, which is grounded in social institutions and involve multiple interests, values, and goals. Because of the value-ladenness of this social process, we may be more successful in drawing the line between good and bad forms of medicalization if we were to use the tools of bioethics (e.g., Parens 2013; Kaczmarek 2019). In this line, Kaczmarek (2019) has developed a promising proposal that departs from Horwitz and Wakefield's substantive account. She proposes to adopt a more pragmatic and ethical approach when assessing medicalization. Her account consists of four guiding questions that are meant to help us identify cases of over-medicalization:

> 1. Has X been rightly recognised as a problem?
> – Does X cause or significantly increase the risk of considerable physical or mental discomfort, suffering, impairments or death?
>
> 2. Does recognising X as a problem not result from unfounded, exaggerated social expectations?
> – Is recognising X as a problem not an example of undue limitation of diversity of individuals for the sake of normalisation? […]
>
> 3. Does medicine provide the most adequate methods of understanding X and its causes?
> – At which level (e.g., molecular, mental, social, several levels combined) do main causes of X occur?
> – Are there any alternative, non-medical and more appropriate ways of understanding X and its causes?
>
> 4. Does medicalizing X ensure the most effective and safest methods of solving it?
> – Are there any alternative, non-medical and more effective ways to solve X or its causes?
> – Does medicalizing X do less harm than good? (Kaczmarek 2019, 122-123)

[6] Note that Horwitz and Wakefield do not deny the importance of social and cultural values in the determination of what a mental disorder is. Rather, they argue that another component plays a role (or should play a role) in the identification of mental disorder: biological dysfunction. This is the claim that I reject here (at least the claim that biological dysfunction is value-free, see Gagné-Julien (forthcoming)). Without this value-neutral component entering into the definition of mental disorder, it is fair to turn to bioethical approaches to assess medicalization.

Identifying a case of over-medicalization would require positive answers to these four questions. While the answers given can be a matter of degree, answering "yes" to all of them means that X has been rightly medicalized. By contrast, answering "no" to all of them would mean that X has been over-medicalized.

I think the four questions and sub-questions identified by Kaczmarek do a good job of covering the issues that are generally associated with the consequences of medicalization mentioned earlier, and reflect the complexity of the medicalization process as well. That is, the four questions appropriately touch on all aspects at stake in the debate on over-medicalization. For instance, the issue of a unilateral understanding of normality and the risk of medicalizing social deviance is well addressed by questions 1 and 2. Question 3 targets the risk involved in depreciating the external causes (social, environmental) of distress. Question 4 refers to the benefits and potential harms of a medical approach for patients. Moreover, I believe that Kaczmarek's account can serve as a good alternative to the substantive approach, in that it does not presuppose any conditions to be "real" medical problems, discovered through a "true" definition of mental disorder. Acknowledging that the characterization of these conditions is a pragmatic task rather than a discovery opens up a space for discussion. It opens a space to discuss each of these issues in acknowledging that giving an answer to these is a pragmatic task, not a discovery. On another note, I believe that, while Kaczmarek's proposal can satisfyingly identify cases of over-medicalization, it could also be used to assess conditions that have not been medicalized yet. That is, despite the fact that the account focuses on over-medicalization, I see no reasons to restrict its use to such cases. For instance, we could use it to assess cases of "under-medicalization", in which people living with a particular condition—which is not currently understood to be medical—would benefit from medicalization (i.e. cases about which we would answer "yes" to most of the four questions). Kaczmarek's proposal could then apply to more cases than simply those which are instances of over-medicalization, and more generally instances of "wrongful medicalization".

Despite the fact that Kaczmarek's contribution is promising, it faces potential problems. First, each of the guiding questions she proposes seems very hard to answer, a problem she acknowledges herself. While Kaczmarek discusses some possible avenues for answers, she does not specify *how* these questions are supposed to be answered and, more importantly, *by whom*. Who is to say, for instance, that seeing X as a problem does not result from an exaggerated social expectation (in response to question 2), or that a non-medical approach would be more effective than a medical one to solve X (in answer to question 4)? Are these answers to

be provided by psychiatrists, bioethicists, patients or citizens?[7] Therefore, even though her account appears to me to be a step in the right direction—because it is not based on a "substantive" conception of mental disorder or on the "true" boundaries of psychiatry—more needs to be said regarding the procedures through which these questions should be answered, and the relevant actors who should express themselves about good and bad forms of medicalization. In the rest of this paper, I will use the EI framework to specify Kaczmarek's pragmatic account. This will lead me to defend an inclusive account of the manner in which the four questions she proposes should be answered.

2. Epistemic Injustices and Problematic Forms of Medicalization

2.1 Epistemic Injustices

I believe that the EI framework as it has been developed by Fricker (2007, 2017; see also e.g., McKinnon 2016; Kidd, Medina, and Pohlhaus 2017) can help us expand and specify Kaczmarek's account, which will allow for a better distinction of good and bad forms of medicalization in psychiatry. This is so because it gives us a better grasp on some forms of injustices that can be created by the process of medicalization, injustices which are often overlooked in the bioethical literature on medicalization. In what follows, I briefly describe the EI framework and state the reasons why it can prove fruitful concerning medicalization in psychiatry. I then present recent work in which these conceptual resources have been applied to medicalization or medicalization-related processes, and show how it could be applied to Kaczmarek's account as well.

EI are wrongs related to the production and transmission of knowledge. The literature generally identifies two types of EI: testimonial injustice and hermeneutical injustice.[8] Testimonial injustice occurs when a hearer deflates the credibility of the speaker because of a negative identity prejudice. In other words, the speaker is not taken seriously by the hearer, not because of her lack of expertise, but because of negative stereotypes related to her belonging to a socially subordinated group (such as in the cases of racism, sexism, classism, etc.—note that these social identities can intersect) (Fricker 2007, 16-17). In the case of testimonial injustice, an epistemic agent is undermined in her capacity to share knowledge. Pre-emptive testimonial injustice is a particular form of testimonial injustice

[7] This point is raised in the debate surrounding the definition of overdiagnosis by Carter et al. (2018). I think it can be applied to Kaczmarek's account as well.

[8] See also e.g., Dotson (2011, 2014) and Berenstain (2016) for more recent work going beyond these two notions.

that occurs when epistemic agents are not solicited in the process of knowledge production, and therefore do not even have the chance to produce their testimony, when such testimony could be relevant. Their testimony is therefore discredited in advance because of a devaluation of the credibility of members of a group which is socially stigmatized or subordinated by the group in power. It is an injustice if their perspective would be relevant to the knowledge-production process, but because of social identity prejudice, it is not even heard (Fricker 2007, 130).

In contrast, hermeneutical injustice happens "when a gap in collective interpretive resources puts someone at an unfair disadvantage when it comes to making sense of their social experiences" (Fricker 2007, 1). In the case of hermeneutical injustice, epistemic agents are wronged in their capacity to understand and/or participate in the collective understanding of the social world. This type of injustice happens to individuals belonging to marginalized social groups, those groups being disadvantaged regarding the availability of or their access to means of creating interpretive resources (e.g., concepts, social schema, etc.) which can make particular aspects of their lived experience intelligible to themselves and others. Testimonial and hermeneutical injustices are injustices because of their discriminatory nature and because of the harmful consequences that they cause to wronged individuals (e.g., loss of confidence as an epistemic agent, feeling of isolation or confusion, etc.).

2.2 Assessing Wrongful Medicalization within an EI Framework

Recall that the main limitation of Kaczmarek's account so far is the vagueness of the procedures through which we are to answer the four suggested questions. Applying EI to her account can prove fruitful for at least two reasons. First, because medicalization is a process of meaning transformation, EI gives us the resources to identify injustices that can happen in relation to this kind of knowledge production. As mentioned earlier, medicalization is the social process through which non-medical phenomena are reinterpreted as medical problems, often as "pathologies" or "disorders". Understood as such, medicalization has an "epistemic tone" (Wardrope 2014). Since it implies the transformation of collective hermeneutic resources to make sense of specific phenomena, here mental distress as a medical problem, and the development of epistemic tools to approach the medicalized conditions,[9] it can be seen as an epistemic process. Therefore, EI could well apply to medicalization and help identify ethical harms that can be created during medicalization, understood as an

[9] Epistemic tools such as concepts, models, and theoretical frameworks (here e.g., the biomedical model of psychiatry, etc.).

epistemic process. This is important for Kaczmarek's account, since answering the four questions—and therefore determining whether a condition should be medicalized or not—is an epistemic process that could create epistemic injustices.

The second reason why EI can prove useful is that it is a good framework to identify injustices that involve social subordination in an epistemic context. For EI to happen, there must be power relations at play: a group is socially subordinated, and such subordination impacts access to knowledge, knowledge creation and/or knowledge transmission. As medicalization scholars have already pointed out, medicalization implies different actors which do not have the same status and level of recognition (here I focus on people living with mental illness versus psychiatrists and psychiatric institutions, see e.g., Reiheld 2010; Wardrope 2014). As Wardrope argues (2014), patients are a marginalized social group during the medicalization process, and medicine (and psychiatry) has excessive power over the construction of conceptual resources related to medicalized phenomena. In other words, medicine has an epistemic privilege regarding the conceptualization of "life problems" (see also Carel and Kidd 2014 for a similar point), while people living with mental illness are underprivileged in that regard. EI can thus help identifying the wrongs associated with social subordination during the medicalization process. So far, because Kaczmarek's account is underspecific about the procedures through which the four questions are to be answered, it cannot keep such power relations from harmfully impacting the medicalization process. But do these EI actually happen during medicalization?

2.3 Hermeneutical and Pre-Emptive Testimonial Injustices Induced by Medicalization

Recent work done in an EI-informed perspective has shown that medicalization can create *hermeneutical* injustices. Fricker has already acknowledged that the medical lexicon and categorization process constrain our collective understanding of what is medically normal and abnormal (Fricker 2007, 163-167). Usually, the hermeneutic resources we draw on to understand phenomena associated with (mental) disorders are forged by medical language. Our collective understanding of mental disorders—because it is developed primarily through psychiatric discourse—masks or dims other dimensions that may be associated with the experience of mental illness. For instance, patients' experiences may be understood only in biomedical terms because of the dominance of hermeneutic resources created by neuro-oriented psychiatry over other, marginalized conceptual models, such as phenomenological approaches (see also Charland 2004, 2013; Conrad and Barker 2010). Wardrope

(2014) explores this further by arguing that medicalization can bring about hermeneutical injustices because patients' experiences are construed *solely* through the discourse of medicine. Because of the power of these medical concepts, patients might not be able to adequately understand what they are experiencing, making it a case of hermeneutical injustice. Despite the occurrence of these epistemic harms in some cases of medicalization, Wardrope adopts a nuanced stance toward the medicalization process. He argues that medicalization can also *provide* hermeneutic resources for patients to report their experiences (for a similar point, see Reiheld 2010). When we look at personal experiences of medicalization, we find that testimonies include a great variety of responses to the process, ranging from positive to negative attitudes (more on this in section 3). Therefore, medicalization in itself does not necessarily create hermeneutical injustices. Only when it deprives patients of access to conceptual resources, or of the means to create hermeneutical tools allowing them to make better sense of their experience, can it be said to create hermeneutical injustices.

Moreover, some recent work by Bueter (2019) on the DSM revision process has revealed a particular form of testimonial injustices. Bueter's analysis does not target the medicalization process itself, but I believe that many aspects of her analysis can fruitfully apply to it. She argues that patients' perspectives are given little consideration when decisions are made about naming conventions, inclusion or exclusion of a condition as a mental disorder, determination of diagnostic thresholds for particular categories, and choices of diagnostic criteria. However, there are good reasons to believe that patient input would be relevant, as in the case of first-person experiences provided by patients about the effects and appropriateness of a particular diagnostic classification (Bueter 2019; see also Carel and Kidd 2014; Scrutton 2017; Drożdżowicz 2021 for patients' particular knowledge and epistemic injustices, but also see Tekin 2020 for the idea of patients' expertise).[10] Patients can provide relevant input regarding how particular conditions are described, and draw attention to overlooked symptoms (Bueter 2019).[11] Patients can also be aware of what is best for them when it comes to the harms and benefits the creation of a

[10] Bueter argues that patients are excluded from the DSM revision process not because they belong to the social group of "patients," but to the social group of "non-experts" (Bueter 2019, 1071). The social identity prejudice at play here would be the negative attitude of experts toward non-experts. While this point is interesting, here I am more interested in epistemic injustices done to patients *qua* belonging to the social group of "patients."

[11] Note that Bueter's argument is in line with the literature about community-based participatory research and, more generally, with situated epistemologies in medical and scientific contexts, even if it has been developed in parallel with them (see e.g., Hill Collins, Harding, Code 2006; Wylie 2014; McHugh 2015; Scheman 2015). That is, marginalized communities can contribute relevant input to knowledge production because their perspective is external to the dominant framework.

new diagnosis might bring about, and report their actual needs concerning the conceptualization of particular conditions. And they can draw attention to the positive value of a "pathological" experience which the medical profession might see only in a negative light (Scrutton 2017). Not considering these forms of knowledge would entail epistemic losses (Drożdżowicz 2021) and create pre-emptive testimonial injustice. Since the two main ways through which medicalization occurs in psychiatry are the creation of a diagnostic category and the modification of diagnostic criteria in the DSM (see section 1.1), it is fair to say that Bueter's analysis can well be applied to the medicalization process. Because patients' perspectives about the DSM revisions are not heard enough, and because their perspectives would be relevant to assess medicalization, patients are wronged as epistemic agents. The fact that the DSM revision process does not provide enough spaces for the inclusion of patients' voices about the creation and modification of psychiatric diagnoses means that medicalization can also create pre-emptive testimonial injustice.

These previous results show that medicalization taking place via the DSM revision structures can create hermeneutical injustices and pre-emptive testimonial injustices. These types of injustices have generally been overlooked in the bioethical literature aiming to distinguish good and bad forms of medicalization. They are nonetheless real injustices that should be avoided, especially since medicalization can be interpreted as an epistemic process. Moreover, the previous analyses imply that the way medicalization occurs in current medical practice and in institutions such as the APA and the DSM revision structures leads to epistemic injustices *usually because people living with mental illness are not heard enough in the process*. The DSM revision process causes EI mainly because patients are excluded from decision-making structures (or plainly not heard enough). Even if the DSM revision process were to adopt Kaczmarek's pragmatic model, it would still need to acknowledge the occurrence of EI during medicalization and the necessity to overcome these harms. While I agree with Kaczmarek's pragmatic proposal and the associated four guiding questions, I think that using an EI framework forces one to advocate that medicalization should be done following an epistemic justice ideal, with the goal of avoiding the creation or perpetuation of epistemic injustices which would impair the epistemic legitimacy of people living with mental illness. Kaczmarek's model has so far proposed no procedures to avoid the epistemic harms actually involved in the medicalization of particular conditions in the DSM.

One way to overcome this deficiency is to argue that—if Kaczmarek's model was implemented in the DSM revision process—answers to the four proposed questions should take patients' voices into account—and take

them seriously. This would make epistemic resources related to medicalization more accessible, and therefore reduce hermeneutical and pre-emptive testimonial injustices. Moreover, even if I believe that Kaczmarek's proposed questions satisfyingly cover the problematic issues related to medicalization which have already been pointed out by medicalization scholars, including patients' perspectives could lead to the realization that other questions need to be asked, specifically where the needs and interests of people experiencing medicalization are concerned. Therefore, I believe that if one is to adopt a pragmatic approach like the one put forward by Kaczmarek, the occurrence of EIs should be taken into account, and mechanisms should be developed to fight them. This would call for the consultation of people living with mental illness on the answers to Kaczmarek's four proposed questions (and even for their assessment of the proposed questions, including the possibility to add more questions or to reformulate existing questions if needed).

In order to reduce the risk of EI, I have argued for the consultation of people living with mental illness in the medicalization process associated with the DSM. This does not amount to the exclusion of psychiatric expertise or of the expertise of other relevant experts in such a decision-making process. The perspectives of patients and of various experts are both relevant on this issue, and the implementation of decision-making structures compatible with diversified views would be ideal. Multiple models exist in the literature on participatory sciences—such as community juries, deliberative opinion polls or consensus conferences following the Danish model, where each member comes from a different perspective and tries to find a viable solution to a controversial issue (see e.g., Fung 2003; Smith 2009; Solomon 2015). The assessment of each of these structures in relation to the ideal of epistemic justice advocated here would require more analysis. But, for now, let us say that inclusive decision-making structures would be a first step toward such an ideal, since they allow for negotiation between divergent views, such as between mental health professionals, other relevant experts and patients. Therefore, arguing in favour of the inclusion of patients' voices in the medicalization process does not entail the exclusion of other types of expertise, but rather makes room for the expertise of patients as well.

3. Problematic Medicalization and PMDD

3.1 A Brief History of the Controversy

To see how rewarding it can be to use EI to expand on Kaczmarek's approach in order to distinguish between good and bad forms of

medicalization, I will use the much-debated case of Premenstrual Dysphoric Disorder (PMDD). My goals in this section are to explain how and why PMDD was included as an official mental disorder in the DSM-5, and to briefly assess this decision in accordance with the conclusion of the previous section. This will draw attention to overlooked epistemic injustices and allow me to suggest possible future improvements.

PMDD has been added in the DSM-5 (APA 2013, 171-175) as an official diagnosis and is now classified as a Depressive Disorder. The main criteria for diagnosing PMDD are "mood lability, irritability, dysphoria, and anxiety symptoms that occur repeatedly during the premenstrual phase of the cycle and remit around the onset of menses or shortly thereafter" (APA 2013, 172). It is also associated with physical symptoms such as breast tenderness, joint or muscle pain and weight gain. The prevalence rate is estimated at between 1,8% and 5,8% among the menstruating women[12] population. Before the introduction of PMDD in the DSM-5, premenstrual psychological distress had already been named in the manual. It was first classified in the DSM-III-R (APA 1987) under the name "Late Luteal Phase Dysphoric Disorder" (LLPDD) and added to Appendix A: "Proposed Diagnostic Categories Needing Further Study." In the DSM-IV-TR (APA 1994), LLPDD was renamed "Premenstrual Dysphoric Disorder" (PMDD) and was included in Appendix B: "Criteria Sets and Axes Provided for Further Study." It could also be diagnosed as "Depressive Disorder Not Otherwise Specified". With the publication of the DSM-5, PMDD was given its full diagnostic status, and was considered to be an official mental disorder (see e.g., Zachar and Kendler 2014 for a more complete history).

The creation of PMDD (and its previous existence as a non-official diagnosis in the DSM) has been criticized from a feminist point of view. The main criticisms concerning PMDD target the illegitimate pathologization and stigmatization of the physical and behavioural changes experienced by women during the premenstrual phase. Moreover, it has been argued that PMDD wrongfully medicalizes the normal distress or anger related to social circumstances such as toxic relationships, history of abuse or social inequalities affecting women (see e.g., Offman 2004;

[12] Note that the DSM and many studies on PMDD refer to "menstruating women" as the only individuals affected by the condition (e.g., APA 2013, 173). However, it should be noted that AFAB (assigned female at birth) individuals can suffer from PMDD. This does not only include cisgender women, but also transgender men, and transmasculine and non-binary individuals. Therefore, when I refer to the way the DSM conceptualizes PMDD, I will use "women" only, and when I talk about PMDD in general, I will use "AFAB individuals" to include cisgender women, transgender men, and transmasculine and non-binary individuals. I take this failure to mention AFAB individuals who are not cisgender women to be a problematic assumption in the DSM's account of the disorder.

Hartlage et al. 2014; Chrisler and Gorman 2015; see also Browne 2015 for a good review).[13] Given the outcry among feminist critics, it might be relevant to investigate the rationale behind the decision to move PMDD to the official list of diagnoses in the DSM-5 in order to assess it.

During the DSM-5 revision process, the Mood Disorders Work Group, in charge of PMDD, mandated a panel of experts specializing in women's mental health to formulate recommendations about PMDD. Epperson and colleagues, members of the panel, published a report in which they explain the reasons motivating the official inclusion of PMDD in the DSM-5. They write that the panel was in charge of

> 1) evaluat[ing] the previous criteria for premenstrual dysphoric disorder, 2) assess[ing] whether there is sufficient empirical evidence to support its inclusion as a diagnostic category, and 3) comment[ing] on whether the previous diagnostic criteria are consistent with the additional data that have become available. (Epperson et al. 2012, 465)

All of the eight members of the panel represented a different country, and six of them were experts of PMDD or reproductive mood disorder. The panel conducted a review of the literature on PMDD. Based on this review and on their discussions, they ultimately recommended that PMDD be moved from the appendix to the Mood Disorders section of the DSM. This decision to include PMDD in the official list of disorders was based on the *Guidelines for Making Changes to DSM-V* produced by Kendler et al. (2009) and used by the different Work Groups assigned to specific revisions. These guidelines are in line with the long-standing wish of the APA to enhance the role of empirical validation in the DSM-5 revision deliberative process (see e.g., Kendler 2013). The document produced by Kendler and colleagues is therefore an overview of qualitative guidelines to advise specific Work Groups in their evaluation of empirical support for proposed modifications to diagnostic categories. It prescribes distinctiveness of diagnosis, and three types of validators: antecedent (e.g., familial aggregation such as family or twin studies), concurrent (e.g., biological markers, patterns of comorbidity) and predictive (e.g., diagnostic stability, course of illness and response to treatment). If a condition meets the validation standards and shows sufficient distinctiveness from other diagnoses, then it can be included in the official nosology.

[13] Note that I cannot do justice to the full and complex history of the controversy surrounding the medicalization of the menstrual cycle. For a more detailed presentation of some of these issues, see e.g. Offman and Kleinplatz (2004), and Chrisler and Caplan (2002).

According to the panel in charge of PMDD, the diagnosis meets all validation requirements. In short, it first appears that PMDD is at least partly heritable. Second, while not associated with a clear biomarker, it appears that the symptoms of PMDD are correlated with menstrual cycle-related hormone fluctuations. Third, PMDD symptoms are generally stable in that they are recurrent at every menstrual cycle (Epperson et al. 2012; Epperson 2013). Moreover, the panel reports that PMDD can be seen as a distinct diagnosis, mainly because of the key correlation between phases of the condition and the menstrual cycle. PMDD seems to be distinct from other diagnoses such as Major Depression (MD) or Bipolar Disorder (BD) since its symptoms are related to the late luteal phase (Epperson et al. 2012, 466-467). Therefore, the main rationale for the inclusion of PMDD as a new diagnosis in the DSM-5 follows the more general empirical turn taken by the DSM during its last revision process, which requires a careful review of empirical evidence to justify the inclusion of new diagnoses.

Nonetheless, in addition to these empirical concerns, it is worth mentioning that the panel reports discussing the feminist worries mentioned earlier concerning the pathologization of women's reproductive cycle and the correlated risk of stigmatization. However, the panel ended up dismissing these worries given the benefits allegedly incurred by the creation of the diagnosis (Epperson et al. 2012, 470; Gotlib and LeMoult 2014). These benefits take into account the decreased functioning of women with PMDD symptoms and include the expected development of therapeutic resources associated with its inclusion in the DSM (Epperson et al. 2012, 470). Studies suggest that the quality of life of women living with severe forms of PMDD were comparable to the one of patients living with MDD (Pearlstein et al. 2000; Halbreich et al. 2003; Rapkin and Winer 2009; Pilver et al. 2013; Osborn et al. 2020a, b). The benefits of including PMDD in the DSM for mental health was held to outweigh the risk of stigmatization and pathologization of feminine anger, especially because the description of the diagnosis made it clear that PMDD concerned only a small minority of women with severe symptoms and could not apply to all women. Thus, despite the fact that there has been no unanimous agreement on the creation of PMDD, it was justified by the panel with arguments about the empirical validity of the disorder and the benefits of this inclusion in terms of future research opportunities and access to clinical care for women with severe symptoms of PMDD.

3.2 Assessing the Medicalization of PMDD in the DSM

I will now turn to the use of Kaczmarek's account and the EI framework to assess the medicalization of PMDD in the DSM-5. I will briefly discuss how the rationale behind the panel's recommendations can be interpreted

as a good fit with Kaczmarek's four questions, but then I will quickly move to the assessment of the creation of PMDD using the tools of EI. I proceed in this manner because I want to focus on how importing an EI framework into Kaczmarek's model can help it shed light on overlooked ethical issues related to the medicalization process in the DSM.

A first thing to note is that the panel in charge of revising the status of PMDD discussed many of the issues covered by Kaczmarek's model. For instance, in discussing the empirical validity of the diagnosis, they addressed question 3 (at least partly), pondering the most adequate methods for understanding a condition and its etiology. For the panel, findings about the empirical validity of the diagnostic category are in favour of its medicalization. Moreover, the panel was concerned with the impact the official inclusion of PMDD in the DSM would have for people living with associated symptoms, especially in terms of access to clinical care. The perceived benefits of the introduction of PMDD as an official diagnosis were seen as an additional argument for its validity—which can be related to questions 1 and 4 (the recognition of a condition as a problem, in terms of suffering or impairment, and the positive effect of medicalization). The risk of harmful pathologization and stigmatization associated with the medicalization of PMDD has also been discussed, in relation with question 4 (Does medicalizing X do less harm than good?). But some sub-questions have also been left unaddressed, such as some sub-questions to question 3, concerning mostly the possible existence of non-medical frameworks to conceptualize and address the condition. Nonetheless, if we interpret the panel's decision within Kaczmarek's framework, it could be argued that the panel asked many of the relevant questions, and that they judged that the medicalization of PMDD would lead to more positive answers than negative ones. Even if the discussions among members of the panel could have gone deeper to address overlooked aspects of medicalization, it could be suggested that including PMDD as an official diagnosis in the DSM is legitimate since Kaczmarek's framework had been applied (recall that this is a matter of degree, and that while medicalizing PMDD can bring about negative consequences, it can still be seen as a legitimate decision given that more questions can be answered by "yes" than by "no").

While it seems that the panel did address many of the core issues of medicalization identified by Kaczmarek, I believe that the PMDD revision process is guilty of creating two types of EI: pre-emptive testimonial injustice and hermeneutical injustice. Looking at the panel's report, AFAB individuals living with PMDD have been left out of the decision-making process. Pre-emptive and hermeneutical injustices occurred because the decision-making process associated with PMDD was not inclusive enough.

As seen in section 2.3, using the EI framework to assess problematic cases of medicalization requires us to make room for consultation and critical discussion involving individuals who will be affected by the process. Within the framework of EI, if individuals with PMDD had been included in the process, and their voices and reports about their lived experience truly heard, epistemic injustices would have been reduced.

Because the consultation with people affected by PMDD did not take place, it is difficult to know precisely what would have been the result of an inclusive process of decision-making grounded in EI. However, recent investigations on PMDD have looked into the narratives of women with specific PMDD symptoms (in contrast with reviews including both PMDD and its milder form, PMS), and studied the impact of this diagnosis on their experience (see e.g., Usher 2014; Hardy and Hardie 2017; Osborn et al. 2020a, b). What these studies reveal is a positive attitude toward the creation of the diagnosis in women living with PMDD. Being diagnosed with PMDD (instead of receiving another diagnosis or no diagnosis at all) was perceived as a relief by most women, who felt that their experience was finally rightfully described:

> I also feel like now I know why, like I know why I feel so anxious sometimes and why I feel so sad. I know it's not my fault, which is probably the main thing, I know it's not my fault now, I'm not just a bad person. (Participant 3) (Reported in Osborn et al. 2020a)

Women diagnosed with PMDD reported feelings of recognition, and of being really heard. They also detailed how the diagnosis transformed their identities and self-understanding, a transformation some described as life-saving. A negative attitude on their part was rather directed toward their "lost years", during which they were not recognized as suffering from PMDD.

As Osborn and colleagues suggest, the positive attitude seen in diagnosed women could be explained in large part by the severe psychological distress associated with PMDD. Participants report:

> All of a sudden it went pitch black, my emotional mood changed drastically and I could never see any outside things, like things had happened that made me upset or made me dark, so as a very young woman I was wondering why I felt that darkness. I felt like there was no point in living.

> I couldn't control the way that I was feeling, I'd cry at the drop of a hat and I'm not particularly a cry, a crying kind of person. It takes quite a lot to get me upset, erm, I just literally could not function. I couldn't, I didn't want to get out of bed in the morning, couldn't sleep at night, erm … just doing stupid things like ripping wallpaper off because I couldn't cope with the anxiety, the feeling of the anxiety. (Reported in Osborn et al. 2020a)

Of course, more research needs to be conducted before we are able to conclude (or overrule) that the medicalization of PMDD is unanimously or mostly welcomed by individuals living with associated symptoms.[14] But these findings suggest that if individuals with PMDD were included in the discussions related to the introduction of PMDD in the DSM-5, they could have asked for its introduction. This would mean that patient requests are in part compatible with the decision of the panel in charge of PMDD.

However, what needs to be emphasized here is that within the EI framework, this does not make PMDD a perfectly good form of medicalization in terms of epistemic justice. This is so because people living with PMDD have not been properly consulted. Despite the fact that patients seem to favour the introduction of PMDD in the DSM-5, their narratives have been collected *after* the inclusion of the diagnosis. During the DSM revision process, these findings were not known. Official structures of consultation and inclusion during the revision process would have *made sure* that the diagnosis as it is described in the DSM meets the needs of people living with PMDD symptoms and matches their interests. It would also have contributed to a more egalitarian access to the creation of hermeneutical resources. One potentially overlooked aspect in these studies is the possibility that, while people living with PMDD symptoms are in need of recognition and care, they might not want their condition to be viewed as a *disorder*. That is, they might want medicalization of PMDD without its pathologization (see e.g., Browne 2015 for a similar point). In another research about PMS more generally, women report hypersensibility to environmental changes and a "deep feeling of vulnerability, a desire to protect themselves from the assaults of everyday life, and of the demands of others; of wanting to turn inwards" (Usher 2014, 318). These types of narratives could help shape the clinical description of PMDD to make sure that people living with associated symptoms recognize themselves in the diagnosis as they would express it,

[14] For instance, only English speaking women over 18 years old who had already received a diagnosis of PMDD were included in Osborn and colleagues' study. But this is a first step toward understanding the attitude of women living with PMDD symptoms toward their diagnosis.

and that the diagnosis is a hermeneutical tool that can really make sense of their experiences. In addition, while it appears clear that most women wish that their symptoms be alleviated, the available treatments tend to focus on medication and, when medication proves ineffective, total hysterectomy combined with bilateral oophorectomy. Women might also want recognition and care, but not necessarily medication or invasive procedures (especially if medication is ineffective for some and if infertility brought on by total hysterectomy is unwanted for many, see Osborn et al. 2020). The treatments developed could be more diversified, and include psychologically based interventions (Usher 2002; see also Usher et al., 2002; Hunter et al., 2002). These are all unexplored possibilities so far. Nonetheless, they point to the epistemic injustices at play in medicalizing PMDD, and to the need for a more inclusive approach to decision-making in the DSM revision process. If such a process were implemented, it would be possible to obtain a medicalized description of PMDD that would reduce epistemic injustices, because it would have been developed in collaboration with people living with PMDD.

Adopting the EI framework shows that it might not be enough to adopt Kaczmarek's pragmatic proposal for identifying good and bad forms of medicalization. The inclusive manner in which the process of medicalization is conducted is relevant to reduce epistemic injustices and to achieve better forms of medicalization. Despite the fact that there is a clear need for recognition and care on the part of people living with PMDD symptoms, further consultation and discussion is needed before we can see PMDD as a fully legitimate form of medicalization. Using the EI framework allows us to pave the way for these possible future improvements.

4. Conclusion

The goal of this paper was to explore the ways in which the EI framework can serve to expand on Kaczmarek's bioethical account, which attempted to distinguish between good and bad instances of medicalization. Kaczmarek's proposal is promising, but it lacks guidance on how the four questions she proposed should be answered, and by whom. Building on the EI framework, I have argued that medicalization in psychiatry can create at least two types of EI: hermeneutical injustice and pre-emptive testimonial injustice. I have then argued that, if Kaczmarek's account was to be implemented, inclusive procedures should be established when debating the medicalization of particular conditions through the DSM in order to address these injustices. This means that individuals living with mental illness should be involved in the discussions and decisions about

the medicalization of their conditions. This is so because medicalization is essentially a process of hermeneutical transformation and comes with power relations between psychiatrists and patients. I have used the controversial case of PMDD to briefly illustrate how using this framework could help make the medicalization of this particular diagnosis more ethical.

What I have proposed here is a first step toward a broader analysis of EI and medicalization in psychiatry. I do not claim to have offered a comprehensive analysis. For instance, a separate analysis drawing on the EI framework would be required to address the role of the pharmaceutical industry as a major driving force of medicalization (e.g., Moynihan and Henry 2002; Moynihan et al. 2013; Musschenga et al. 2010). Moreover, recent work suggests that EI can also occur among patient advocacy groups (Jongsma et al. 2017; Jordan et al. 2020; Matthew et al. 2020), raising the question of how to prevent EI coming from patients' organizations themselves.

In addition, I will signal several questions which I have left unanswered in this paper: How to ensure that patients' voices are truly heard in an ethical medicalization process? How should critical discussions with patients be conducted? And how to deal with serious disagreement between participants (e.g., between patients and psychiatrists, or between patients)? What this list of questions suggests is that research needs to be urged further in order to better map the many power relations at play in the process of medicalization and the exact ways EI can occur in the DSM revision process. Nonetheless, I do believe that more interaction is required between EI literature and the research on wrongful medicalization. I hope I have been able to contribute to this nascent dialogue.

Acknowledgments

I wish to thank Phoebe Friesen, Amandine Catala, the members of the Canada Research Chair on Epistemic Injustice and Agency and two anonymous referees for their help and comments with previous versions of this paper.

REFERENCES

Bandini, Julia. 2015. "The Medicalization of Bereavement: (Ab)normal Grief in the DSM-5." *Death Studies* 39: 347–352. https://doi.org/10.1080/07481187.2014.951498.

Batstra, Laura, and Allen Frances. 2012. "Holding the Line Against Diagnostic Inflation in Psychiatry." *Psychotherapy and psychosomatics* 81 (1): 5-10. https://doi.org/10.1159/000331565.

Berenstain, Nora. 2016. "Epistemic Exploitation." *Ergo* 3 (22). http://doi.org/10.3998/ergo.12405314.0003.022.

Bingham, Rachel and Natalie Banner. 2014. "The Definition of Mental Disorder: Evolving but Dysfunctional?" *Journal of Medical Ethics* 40 (8): 537-542. http://dx.doi.org/10.1136/medethics-2013-101661.

Browne, Tamara Kayali. 2015. "Is Premenstrual Dysphoric Disorder Really a Disorder?" *Journal of Bioethical Inquiry* 12 (2): 313-330. https://doi.org/10.1007/s11673-014-9567-7.

Browne, Tamara Kayali. 2017. "A Role for Philosophers, Sociologists and Bioethicists in Revising the DSM: A Philosophical Case Conference." *Philosophy, Psychiatry, & Psychology* 24 (3): 187-201. https://doi.org/10.1353/ppp.2017.0024.

Bueter, Anke. 2019. "Epistemic Injustice and Psychiatric Classification." *Philosophy of Science* 86 (5): 1064-1074. https://doi.org/10.1086/705443.

Busfield, Joan. 2017. "The Concept of Medicalisation Reassessed." *Sociology of Health & Illness 39(5):* 759–774. https://doi.org/10.1111/1467-9566.12538.

Carel, Havi, and Ian James Kidd. 2014. "Epistemic Injustice in Healthcare: A Philosophical Analysis." *Medicine, Healthcare, and Philosophy* 17 (4): 529-540. https://doi.org/10.1007/s11019-014-9560-2.

Carter, Stacy M., Rogers, Wendy, Heath, I, Degeling, Chris, Doust, Jenny, and Alexandra Barratt. 2015. "The Challenge of Overdiagnosis Begins with Its Definition." *British Medical Journal* 350. https://doi.org/10.1136/bmj.h869.

Carter, Stacy. M., Degeling, Chris, Doust, Jenny and Alexandra Barratt. 2016. "A Definition and Ethical Evaluation of Overdiagnosis." *Journal of Medical Ethics* 42 (11): 705-714. http://dx.doi.org/10.1136/medethics-2015-102928.

Crichton, Paul, Carel, Havi, and Ian James Kidd. 2017. "Epistemic Injustice in Psychiatry." *British Journal of Psychiatric Bulletin* 41 (2): 65-70. https://doi.org/10.1192/pb.bp.115.050682.

Charland, Louis. 2013. "Why Psychiatry Should Fear Medicalization?" In *The Oxford Handbook of Philosophy and Psychiatry*, edited by K.W.M. Fulford, Martin Davies, Richard G.T. Gipps, George

Graham, John Z. Sadler, Giovanni Stanghellini and Tim Thornton, 159-175. Oxford: Oxford University Press.

Chrisler, Joan C. and Paula Caplan. 2002. "The Strange Case of Dr. Jekyll and Ms. Hyde: How PMS Became a Cultural Phenomenon and a Psychiatric Disorder." *Annual Review of Sex Research* 13 (1): 274-306. https://doi.org/10.1080/10532528.2002.10559807.

Conrad, Peter. 2007. *The Medicalization of Society: On the Transformation of Human Condition into Treatable Disorders*. Baltimore: Johns Hopkins University Press.

———. 2013. "Medicalization: Changing Contours, Characteristics, and Contexts." *In Medical Sociology on the Move: New Directions in Theory,* edited by William C. Cockerham, 195-214. Dordrecht*:* Springer.

Conrad, Peter and Kristin K. Barker. 2010. "The Social Construction of Illness: Key Insights and Policy Implications." *Journal of Health and Social Behavior* 51 (1): 67-69. https://doi.org/10.1177/0022146510383495.

Conrad, Peter. Mackie, Thomas and Ateev Mehrotra. 2010. "Estimating the Costs of Medicalization." *Social Science & Medicine* 70 (12): 1943-1947. https://doi.org/10.1016/j.socscimed.2010.02.019.

Conrad Peter and Caitlin Slodden. 2013. "The Medicalization of Mental Disorder." In *Handbook of the Sociology of Mental Health,* edited by Carol S. Aneshensel, Jo C. Phelan and Alex Bierman, 61-73. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-4276-5_4.

Davis, Joseph. E. 2010. "Medicalization, Social Control, and the Relief of Suffering." In *The New Blackwell Companion to Medical Sociology,* edited by William C. Cockerham, 211-241. Malden, MA: Wiley-Blackwell.

Davis, Emmalon. 2018. "On Epistemic Appropriation." *Ethics* 128 (4): 702-727. https://doi.org/10.1086/697490.

Dotson, Kristie. 2011. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia: A Journal of Feminist Philosophy* 26 (2): 236–257. https://doi.org/10.1111/j.1527-2001.2011.01177.x.

———. 2014. "Conceptualizing Epistemic Oppression." *Social Epistemology* 28 (2): 115–138. https://doi.org/10.1080/02691728.2013.782585.

Drożdżowicz, Anna. 2021. "Epistemic Injustice in Psychiatric Practice: Epistemic Duties and the Phenomenological Approach." *Journal of Medical Ethics.* http://dx.doi.org/10.1136/medethics-2020-106679.

Epperson, Neil C. 2013. "Premenstrual Dysphoric Disorder and the Brain." *The American journal of psychiatry* 170 (3).

https://doi.org/10.1176/appi.ajp.2012.12121555.

Epperson, Neil C., Steiner, Meir, Hartlage, S. Ann., Eriksson, Elias, Schmidt, Peter. J., Jones, Ian and Kimberly A. Yonkers. 2012. "Premenstrual Dysphoric Disorder: Evidence for a New Category for DSM-5." *American Journal of Psychiatry* 169 (5): 465-475. https://doi.org/10.1176/appi.ajp.2012.11081302.

Faucher, Luc. 2021. "Facts, Facts, Facts: HD Analysis Goes Factual." In *Defining Mental Disorders: Jerome Wakefield and his critics,* edited by Luc Faucher and Denis Forest, 47-70. Cambridge, MA: MIT Press.

Frances Allen. 2010, March 1. "It's Not Too Late to Save 'Normal'." *LA Times.* Retrieved from: http://articles.latimes.com/

———. 2013. *Saving Normal: An Insider's Revolt Against Out-Of-Control Psychiatric Diagnosis, DSM-5, Big Pharma and The Medicalization of Ordinary Life*. New York, N.Y.: William Morrow.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford: Oxford University Press.

———. 2017. "Evolving Concepts of Epistemic Injustice." In *The Routledge Handbook of Epistemic Injustice,* edited by Ian James Kidd, José Medina and Gaile Pohlhaus Jr, 53-60. New York: Routledge.

Fung, Archon. 2003. "Recipes for Public Spheres: Eight Institutional Design Choices and Their Consequences." *Journal of Political Philosophy* 11: 338–367. https://doi.org/10.1111/1467-9760.00181.

Gagné-Julien, Anne-Marie. Forthcoming. "Dysfunction and the Definition of Mental Disorder in the DSM." *Philosophy, Psychiatry & Psychology*.

Gosselin, Abigail. 2018. "Mental Illness Stigma and Epistemic Credibility." *Social Philosophy Today* 34: 77-94. https://doi.org/10.5840/socphiltoday20185852.

———. 2019. ""Clinician Knows Best"? Injustices in the Medicalization of Mental Illness." *Feminist Philosophy Quarterly* 5 (2). https://doi.org/10.5206/fpq/2019.2.7285.

Gotlib, Ian H. and Joelle LeMoult. 2014. "The "Ins" and "Outs" of the Depressive Disorders Section of DSM-5." *Clinical Psychology: Science and Practice* 21 (3): 193-207. https://doi.org/10.1111/cpsp.12072.

Halbreich Uriel, Borenstein Jeff, Pearlstein Terry and Linda S. Kahn. 2003. "The Prevalence, Impairment, Impact, and Burden of Premenstrual Dysphoric Disorder (PMS/PMDD)." *Psychoneuroendocrinology* 28 (3): 1–23. https://doi.org/10.1016/S0306-4530(03)00098-2.

Hardy, Claire and Jenna Hardie. 2017. "Exploring Premenstrual Dysphoric Disorder (PMDD) in the Work Context: A Qualitative Study." *Journal of Psychosomatic Obstetrics and Gynecology* 38 (4): 292–300. https://doi.org/10.1080/0167482X.2017.1286473.

Hartlage S., Ann, Breaux Cynthia A., and Kimberly A. Yonkers. 2014. "Addressing Concerns about the Inclusion of Premenstrual Dysphoric Disorder in DSM-5." *Journal of Clinical Psychiatry* 75 (1):70–76. https://doi.org/10.4088/JCP.13cs08368.

Healy, David. 2008. *Mania: A Short History of Bipolar Disorder*. Baltimore: John Hopkins University Press.

Hofmann, Bjørn. 2016. "Medicalization and Overdiagnosis: Different but Alike". *Medicine, Health Care and Philosophy*, 19 (2): 253-264. https://doi.org/10.1007/s11019-016-9693-6.

Horwitz, Allan V. and Jerome C. Wakefield. 2007. *The Loss of Sadness: How Psychiatry Transformed Normal Sorrow into Depressive Disorder*. Oxford: Oxford University Press.

———. 2012. *All We Have to Fear: Psychiatry's Transformation of Natural Anxieties into Mental Disorders*. Oxford: Oxford University Press.

Hunter, Myra, Ussher, Jane, M., Cariss, Margaret, Browne, Susan, and Rosanne, Jelly. 2002. "A Randomised Comparison of Psychological (Cognitive Behaviour Therapy, CBT), Medical (fluoxetine) and Combined (CBT and Fluoxetine) Treatment for Women with Premenstrual Dysphoric Disorder." *Journal of Obstetrics and Gynaecology* 23 (3): 193-199. https://doi.org/10.3109/01674820209074672.

Jones, Nev, and Robyn Brown. 2012. "The Absence of Psychiatric C/S/X Perspectives in Academic Discourse: Consequences and Implications." *Disability Studies Quarterly* 33 (1). http://dx.doi.org/10.18061/dsq.v33i1.3433.

Jongsma, Karin, Elisabeth Spaeth, and Silke Schicktenz. 2017. "Epistemic Injustice in Dementia and Autism Patient Organisations: An Empirical Analysis". *AJOB Empirical Bioethics,* 8 (4): 28-30. https://doi.org/10.1080/23294515.2017.1402833.

Kaczmarek, Emlia. 2019. "How to Distinguish Medicalization from Over-Medicalization?" *Medicine, Health Care and Philosophy* 22 (1): 119-128. https://doi.org/10.1007/s11019-018-9850-1.

Kidd, Ian James and Havi Carel. 2017. "Epistemic Injustice and Illness". *Journal of Applied Philosophy* 33 (2): 172-190. https://doi.org/10.1111/japp.12172.

———. 2018. "Naturalism, Healthcare Practice, and Epistemic Injustice". *Royal Institute of Philosophy Supplement,* 84: 1-23.

———. Forthcoming. "Pathocentric Hermeneutical Injustice and Conceptions of Health". In *Overcoming Epistemic Injustice: Social and Psychological Perspectives,* edited by Ben Sherman and Stacey Goguin, New York: Rowman and Littlefield.

Kidd, Ian James, Medina, José and Gaile Pohlhaus Jr. (eds). 2017. *The Routledge Handbook of Epistemic Injustice*. London, NY: Routledge.

Kirk, Stuart A. and Herb Kutchins. 1992. *The Selling of DSM: The Rhetoric of Science in Psychiatry*. Hawthorne: Aldine de Gruyter.

Kurs, Rena and Alexander Grinshpoon. 2018. "Vulnerability of Individuals with Mental Disorders to Epistemic Injustice in both Clinical and Social Domains." *Ethics & Behavior* 28 (4): 336-346. https://doi.org/10.1080/10508422.2017.1365302.

Kyratsous, Michalis and Abdi Sanati. 2017. "Epistemic Injustice and Responsibility in Borderline Personality Disorder." *Journal of Evaluation in Clinical Practice* 23 (5): 974-980. https://doi.org/10.1111/jep.12609.

Lane, Christopher. 2007. *Shyness: How Normal Behavior Became a Sickness*. New Haven, CT: Yale University Press.

Lilienfeld, Scott O., and Lori Marino. 1995. "Mental Disorder as a Roschian Concept: A Critique of Wakefield's "Harmful Dysfunction" Analysis." *Journal of Abnormal Psychology* 104 (3): 411-420. https://doi.org/10.1037/0021-843X.104.3.411.

Leblanc, Stephanie, and Elizabeth Anne Kinsella. 2016. "Toward Epistemic Justice: A Critically Reflexive Examination of 'Sanism' and Implications for Knowledge Generation." *Studies in Social Justice* 10 (1): 59-78. https://doi.org/10.26522/ssj.v10i1.1324.

Mason, Rebecca. 2011. "Two Kinds of Unknowing." *Hypatia* 26 (2): 294–307. https://doi.org/10.1111/j.1527-2001.2011.01175.x.

McCoy, Matthew S., Liu, Emily Y., Lutz Amy S. F., and Dominic Sisti. 2020. "Ethical Advocacy across the Autism Spectrum: Beyond Partial Representation." *The American Journal of Bioethics*, 20 (4): 13-24. https://doi.org/10.1080/15265161.2020.1730482.

McNally, Richard J. 2001. "On Wakefield's Harmful Dysfunction Analysis of Mental Disorder." *Behaviour Research and Therapy*, 39 (3): 309-314. https://doi.org/10.1016/S0005-7967(00)00068-1.

Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppressions, Epistemic Injustice, and Social Imaginations*, Oxford: Oxford University Press.

Moynihan Ray, Heath Iona, and David Henry. 2002. "Selling Sickness: The Pharmaceutical Industry and Disease Mongering." *British Medical Journal* 324: 886-890. https://doi.org/10.1136/bmj.324.7342.886.

Moynihan Ray N., Cooke Georga P., Doust Jenny A., Bero Lisa, Hill Suzanne and Paul Glasziou. 2013. "Expanding Disease Definitions in Guidelines and Expert Panel Ties to Industry: A Crosssectional Study of Common Conditions in the United States." *PLoS Med* 10 (8). https://doi.org/10.1371/journal.pmed.1001500

Murano, Maria Cristina. 2018. "Medicalising Short Children with Growth Hormone? Ethical Considerations of the Underlying Sociocultural Aspects." *Medicine, Health Care and Philosophy* 21 (2): 243-253. https://doi.org/10.1007/s11019-017-9798-6.

Murphy-Hollies, Kathleen. 2021. "When a Hybrid Account of Disorder Is Not Enough: The Case of Gender Dysphoria." *European Journal of Analytic Philosophy* 17 (2): (SI6)5-37. https://doi.org/10.31820/ejap.17.3.5.

Murphy, Dominic and Robert L. Woolfolk. 2000. "Conceptual Analysis Versus Scientific Understanding: An Assessment of Wakefield's Folk Psychiatry." *Philosophy, Psychiatry, & Psychology* 7 (4): 271-293.

Musschenga, A.W., van der Steen, W.J. and V.K.Y. Ho. 2010. "The Business of Drug Research: A Mixed Blessing." In *The Commodification of Academic Research*, edited by Hans Radder, 110-131. Pittsburgh: Pittsburgh University Press.

Newbigging, Karen and Julie Ridley. 2018. "Epistemic Struggles: The Role of Advocacy in Promoting Epistemic Justice and Rights in Mental Health." *Social Science & Medicine* 219: 36-44. https://doi.org/10.1016/j.socscimed.2018.10.003.

Offman, Alia and Peggy J., Kleinplatz. 2004. "Does PMDD Belong in the DSM? Challenging the Medicalization of Women's Bodies." *The Canadian Journal of Human Sexuality* 13 (1): 17–27.

Osborn, Elizabeth, Wittkowski, Anja, Brooks, Joanna, Briggs, Paula E. and Saughn P. O'Brien. 2020a. "Women's Experiences of Receiving a Diagnosis of Premenstrual Dysphoric Disorder: A Qualitative Investigation." *BMC Women's Health* 242 (20). https://doi.org/10.1186/s12905-020-01100-8.

Osborn, Elizabeth, Brooks, Joanna, O'Brien, Saughn P., and Anja Wittkowski. 2020b. "Suicidality in Women with Premenstrual Dysphoric Disorder: A Systematic Literature Review." *Archives of Women's Mental Health 24:*173-184. https://doi.org/10.1007/s00737-020-01054-8.

Parens, Erik. 2013. "On Good and Bad Forms of Medicalization." *Bioethics*, 27 (1): 28-35. https://doi.org/10.1111/j.1467-8519.2011.01885.x.

Pearlstein T.B., Halbreich U., Batzar E.D., Brown C.S., Endicott J., Frank E., et al. 2000. "Psychosocial Functioning in Women with Premenstrual Dysphoric Disorder Before and After Treatment with Sertraline or Placebo." *Journal of Clinical Psychiatry* 61 (2):101–109. https://doi.org/10.4088/jcp.v61n0205.

Pies, Ronal. 2014. "The Bereavement Exclusion and DSM-5: An Update and Commentary." *Innovations in clinical neuroscience* 11 (7-8): 19-22.

Pilver, Corey E., Libby, Daniel J., and Rani A. Hoff. 2013. "Premenstrual Dysphoric Disorder as a Correlate of Suicidal Ideation, Plans, and Attempts among a Nationally Representative Sample." *Social Psychiatry and Psychiatric Epidemiology* 48 (3):437–46. https://doi.org/10.1007/s00127-012-0548-z.

Powell, Russell and Eric Scarffe. 2019a. "Rethinking "Disease": A Fresh Diagnosis and a New Philosophical Treatment." *Journal of Medical Ethics*, *45*(9): 579-588. http://dx.doi.org/10.1136/medethics-2019-105465.

———. 2019b. Rehabilitating "Disease": Function, Value, and Objectivity in Medicine. *Philosophy of Science* 86 (5): 1168-1178. https://doi.org/10.1086/705520.

Purdy, Laua 2001. "Medicalization, Medical Necessity, and Feminist Medicine." *Bioethics* 15 (3): 248-261. https://doi.org/10.1111/1467-8519.00235.

Rapkin Andrea J. and Sharon A. Winer. 2009. "Premenstrual Syndrome and Premenstrual Dysphoric Disorder: Quality of Life and Burden of Illness." *Expert Review of Pharmacoeconomics & Outcomes Research* 9 (2): 157–170. https://doi.org/10.1586/erp.09.14.

Reiheld, Alison. 2010. "Patient Complains of…: How Medicalization Mediates Power and Justice." *International Journal of Feminist Approaches to Bioethics* 3 (1): 72-98. https://doi.org/10.3138/ijfab.3.1.72.

Richardson, Jordan P., and Richard R. Sharp. 2020. "Meaningful Fissures: The Value of Divergent Agendas in Patient Advocacy." *The American Journal of Bioethics* 20 (4): 1-3. https://doi.org/10.1080/15265161.2020.1735873.

Riska, Elianne. 2015. "Resisting and Endorsing Medicalization." In A*ging Men, Masculinities and Modern Medicine,* edited by Antje Kampf, Barbara L. Marshall, Alan Petersen, 71-85. London: Routledge.

Rogers, Wendy A. and Yishai Mintzker. 2016. "Getting Clearer on Overdiagnosis." *Journal of Evaluation in Clinical Practice* 22 (4): 580–587. https://doi.org/10.1111/jep.12556.

Rogers, Wendy A. and Yishai Mintzker. 2016. "Getting Clearer on Overdiagnosis." *Journal of Evaluation in Clinical Practice* 22 (4): 580–587. https://doi.org/10.1111/jep.12556.

Sadler, John Z., Jotterand, Fabrice, Lee, Craddock, Simon and Stephen Inrig. 2009. "Can Medicalization be Good? Situating Medicalization within Bioethics." *Theoretical Medicine and Bioethics* 30 (6): 411-425. https://doi.org/10.1007/s11017-009-9122-4.

Sanati, Abdi and Michalis Kyratsous. 2015. "Epistemic Injustice in Assessment of Delusions." *Journal of Evaluation in Clinical Practice*, 21 (3): 479-485. https://doi.org/10.1111/jep.12347.

Schramme, Thomas. 2010. "Can We Define Mental Disorder by Using the Criterion of Mental Dysfunction?" *Theoretical Medicine and Bioethics*, 31 (1): 35-47. https://doi.org/10.1007/s11017-010-9136-y.

Scott, Wilbur J. 1990. "PTSD in *DSM-III*: A Case in the Politics of Diagnosis and Disease." *Social Problems,* 37: 294-310. https://doi.org/10.2307/800744.

Scrutton, Anastasia Philipa. 2017. "Epistemic Injustice and Mental Illness" In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., 347-355. New York: Routledge.

Sedler, Mark. J. 2016. "Medicalization in Psychiatry: The Medical Model, Descriptive Diagnosis, and Lost Knowledge." *Medicine, Health Care and Philosophy* 19 (2): 247-252. https://doi.org/10.1007/s11019-015-9670-5.

Smith, Graham. 2009. *Democratic Innovations: Designing Institutions for Citizen Participation*. Cambridge: Cambridge University Press.

Solomon, Miriam. 2015. *Making Medical Knowledge*. Oxford: Oxford University Press.

Stegenga, Jacob. 2021. "Medicalization of Sexual Desire." *European Journal of Analytic Philosophy* 17 (2): (SI5)5-32. https://doi.org/10.31820/ejap.17.3.4.

Stein, Dan J., Kaminer, Debra, Zungu-Dirwayi Nompumelelo, and Soraya Seedat. 2006. "Pros and Cons of Medicalization: The Example of Trauma." *The World Journal of Biological Psychiatry* 7 (1), 2–4. https://doi.org/10.1080/15622970500483110.

Sullivan, Patrick. 2019. "Epistemic Injustice and Self-Injury: A Concept with Clinical Implications." *Philosophy, Psychiatry, & Psychology* 26 (4): 349-362. https://doi.org/10.1353/ppp.2019.0049.

Tate, Alex James Miller. 2019. "Contributory Injustice in Psychiatry." *Journal of Medical Ethics* 45 (2): 97-100. http://dx.doi.org/10.1136/medethics-2018-104761.

Tekin, Şerife. 2020. "Patients as Experienced-Based Experts in Psychiatry: Insights from the Natural Method." In *The Natural Method: Ethics, Mind & Self, Themes from the Work of Owen Flanagan*, edited by Eddy Nahmias, Thomas W. Polger, Wenqing Zhao, 79-98, Cambridge, MA: MIT Press.

Thomas, Felicity. 2021 "Medicalisation." In *Routledge International Handbook of Critical Issues in Health and Illness*, edited by Kerry Chamberlain, Antonia Lyons, 23-33. London: Routledge.

Ussher, Jane M. 2002. "Processes of Appraisal and Coping in the Development and Maintenance of Premenstrual Dysphoric Disorder." *Journal of Community & Applied Social Psychology* 12 (5), 309-322. https://doi.org/10.1002/casp.685.

Ussher Jane M., Hunter Myra, and Browne Susanna J. 2000. "Good, Bad or Dangerous to Know: Representations of Femininity in Narrative Accounts of PMS." In *Culture in Psychology*, edited by Corinne Squire, 87-99. New York, NY: Routledge.

Ussher, Jane M., Hunter, Myra, and Margaret Cariss. 2002. "A Woman Centred Psychological Intervention for Premenstrual Symptoms, Drawing on Cognitive Behavioural and Narrative Therapy." *Journal of Clinical Psychology and Psychotherapy* 9 (5): 319-331. https://doi.org/10.1002/cpp.340.

Wakefield, Jerome. C. 1992. "The Concept of Mental Disorder: On the Boundary between Biological Facts and Social Values." *American Psychologist* 47 (3): 373-388. https://doi.org/10.1037/0003-066X.47.3.373.

Wakefield, Jerome. C. and Jordan A. Conrad. 2019. "Does the Harm Component of the Harmful Dysfunction Analysis Need Rethinking?: Reply to Powell and Scarffe." *Journal of Medical Ethics* 45 (9): 594-596. http://dx.doi.org/10.1136/medethics-2019-105578.

Wardrope, Alistair. 2015. "Medicalization and Epistemic Injustice." *Medicine, Health Care and Philosophy* 18 (3): 341-352. https://doi.org/10.1007/s11019-014-9608-3.

Zachar, Peter and Kenneth S. Kendler. 2014. "A Diagnostic and Statistical Manual of Mental Disorders History of Premenstrual Dysphoric Disorder." *The Journal of Nervous and Mental Disease* 202 (4): 346-352. https://doi.org/10.1097/NMD.0000000000000128.

European Journal of Analytic Philosophy EuJAP | Vol. 17 | No. 2 | 2021
https://doi.org/10.31820/ejap.17.3.4

MEDICALIZATION OF SEXUAL DESIRE

Jacob Stegenga[1]

[1] University of Cambridge

Original scientific article – Received: 21/05/2021 Accepted: 08/11/2021

ABSTRACT

Medicalisation is a social phenomenon in which conditions that were once under legal, religious, personal or other jurisdictions are brought into the domain of medical authority. Low sexual desire in females has been medicalised, pathologised as a disease, and intervened upon with a range of pharmaceuticals. There are two polarised positions on the medicalisation of low female sexual desire: I call these the mainstream view and the critical view. I assess the central arguments for both positions. Dividing the two positions are opposing models of the aetiology of low female sexual desire. I conclude by suggesting that the balance of arguments supports a modest defence of the critical view regarding the medicalisation of low female sexual desire.

Keywords: *medicalization; female sexual interest/arousal disorder; philosophy of medicine; disease; controversial diseases; philosophy of psychiatry*

1. Introduction

Medicalisation is a social phenomenon in which conditions that were once under legal, religious, personal or other jurisdictions are brought into the domain of medical authority. Low sexual desire in females has been medicalised, pathologised as a disease, and intervened upon with a range of pharmaceuticals. There are two polarised positions on the medicalisation of low female sexual desire. The mainstream view—implicitly held or explicitly articulated by many physicians, patient advocacy groups, pharmaceutical companies, activists, and policy makers—is that the medicalisation of low female sex desire is appropriate. Many females with low sexual desire suffer distress, on the mainstream

Correspondence: jms303@cam.ac.uk

view, and medicine is the correct jurisdiction for the alleviation of such suffering. Sexual desire, on this view, is like an appetite—a function of biological features such as hormone balances or neurotransmitter concentrations—and low sexual desire can be modulated by exogenous interventions on these biological features.

The critical view—implicitly held or explicitly articulated by some psychiatrists, psychologists, journalists, activists, and academic commentators—is that the medicalisation of low female sexual desire is pernicious. These critics argue that low sexual desire ought to be understood not as a disease but rather as a phenomenon arising out of a particular social context, and thus medicine is not the correct jurisdiction for females who experience low sexual desire. Sexual desire, on the critical view, is not solely or typically a function of biological causes but rather is typically a function of social causes—perhaps as a result of stress or fatigue or uneducated partners or toxic relationships or other diseases or even as a harmful effect of medications for those other diseases. Such critics sometimes claim that the very notion that one's sexual desires are dysfunctionally low involves appealing to culturally-determined norms of sexuality, or relational imbalances between the sexual desires of a female and her partner, and are not necessarily intrinsic harms to a female with low desire herself.

In short, there exist two antagonist positions regarding the medicalisation of low female sexual desire. In practice the positions are not always so clearly demarcated—the psychiatrist Rosemary Basson, for example, contributed to the development of the contemporary diagnostic category of low female sexual desire while also criticising the use of pharmaceutical interventions for the alleged disease. Nevertheless, there are clear trenches on the ground, and both sides are armed with statistics, science, patient testimonies, campaigns, and principled arguments of varying quality.

When asked about the potentially nefarious consequences of medicalising low female sexual, Irwin Goldstein, a urologist and prominent defender of the medicalisation of female sexual desire, deflected the concern by responding "that's a question for some philosopher" (Quoted in Moynihan 2003). Here I describe and assess several of the most important arguments from both positions regarding the medicalisation of low female sexual desire.[1] I begin by tracing conceptualisations of low female sexual desire beginning in the early twentieth century (§2). This is stage-setting. I

[1] In this paper I use the term 'female'; although the scientific literature that this paper addresses often uses the terms 'woman' and 'female' interchangeably, the putative disease in question targets the biological category 'female' (and this term appears in the name of the disease), and an inclusion criterion for the clinical studies is status as a biological female.

proceed to articulate and assess several of the most important arguments for the mainstream view (§3) and the critical view (§4). Dividing the two positions are opposing models of the aetiology of low female sexual desire (§5). I conclude by suggesting that the balance of arguments supports a modest defence of the critical view regarding the medicalisation of low female sexual desire (§6).

2. Conceptualizations of Low Female Sexual Desire

Though Foucault flagged the middle of the nineteenth century as the moment in which a sub-discipline of medicine devoted to sex appeared, the focus during this nascent period of sex medicine was the 'paraphilias' or 'sexual perversions' (sexual desire for an atypical object or activity in which such desire causes distress to the desirer or harm to others).[2] Low sexual desire in females has been pathologized by psychiatry and related disciplines since the final years of the nineteenth century (Angel 2010). Marital advice manuals, psychoanalytic texts, psychiatric diagnostic manuals, sexologists, and feminist critics of much of this discourse have articulated numerous theories about low female sexual desire, including what constitutes female sexual dysfunction, and its causes and optimal modes of treatment. There are two broad classes of models of low female sexual desire: an appetitive or biological model, which holds that low female sexual desire is a result of a dysfunction in a physiological capacity, and a social or contextual model, which holds that low female sexual desire is a result of features of a female's social or cultural context (§5).

The way in which low female sexual desire has been conceived has changed often, as illustrated by the various editions of the DSM. The first edition, published in 1952, included 'frigidity', which was the closest of the female sexual dysfunctions in this edition to what we would now call low sexual desire—frigidity was characterised as disinterest in heterosexual intercourse or lack of pleasure from intercourse (other female sexual dysfunctions in the first edition included 'involutional melancholia', dyspareunia, and 'nymphomania'). After the sexual revolution of the 1960s and 1970s, the diagnosis of too much desire (nymphomania) was eliminated from the third edition, published in 1980. The third edition added the category 'inhibited sexual desire' as the diagnosis for low sexual desire in both males and females. The revision to

[2] The Russian physician Heinrich Kaan published his 'Psychopathia Sexualis' in 1846, in which he re-interpreted Christian sins into medical diseases; he characterised masturbation and fantasies to be the basis sexual disorders. In Foucault's 1974-75 lectures at College de France he noted that Kaan's book "was the first treatise of psychiatry to speak only of sexual pathology but the last to speak of sexuality solely in Latin". Kraft-Ebbing's more influential book of the same title appeared forty years later.

the third edition, published in 1987, perhaps cleansing itself of its psychoanalytic hangover, renamed inhibited sexual desire as 'hypoactive sexual desire disorder' (again for both males and females). The present edition of the DSM is the fifth, published in 2013. Hypoactive sexual desire disorder has been divided into a male version (male hypoactive sexual desire disorder), and a female version: female sexual interest/arousal disorder.

Parallel to the evolution of the DSM, developments in the scientific and feminist study of sex provided new ways of conceiving of disorders of sexual desire. From Freud's psychoanalysis and Kinsey's statistics, from the laboratory work of Masters and Johnson, from feminist-inspired sociological, psychological and psychiatric work of those such as Hite and Tiefer and Basson, we now have multiple conceptualisations of the causes and constituents of low female sexual desire.

Freud developed psychoanalysis in part based on the idea that many of our psychopathologies are based on forms of psychological repression, and he most prominently applied this to sex. The frigidity of some women, according to Freud, was a result of psychogenic causes. Famously, Freud (1905) claimed that clitoral orgasms are a sign of immature sexual development, which held some sway into the middle of the twentieth century. Kinsey was critical of the psychoanalytic approach to sexual desire, and instead adopted a 'capacity' model, which held that different people had differing intrinsic sexual capacities. These capacities were physiological in nature, and they manifest in behaviour, specifically the frequency of a person's sexual activities. Females, on average, had lower sexual capacities than males, claimed Kinsey. Kinsey thought that such variability in a physiological sex capacity better explained variability in sexual desires compared with a repression model.[3] Thus Kinsey foreshadowed a disease model of low sexual desire.

This approach was continued by the laboratory studies of Masters and Johnson. They observed people having sexual intercourse and masturbating, and ultimately recorded over ten thousand orgasms while measuring various physiological features, which formed the empirical basis of their four-phase 'sexual response cycle': excitement, plateau, orgasm, and resolution. This theory was influential; for example, it was

[3] Kinsey wrote: "There is an inclination among psychiatrists to consider all unresponding individuals as inhibited, and there is a certain skepticism in the profession of the existence of people who are basically low in capacity to respond. This amounts to asserting that all people are more or less equal in their sexual endowments, and ignores the existence of individual variation. No one who knows how remarkably different individuals may be in morphology, in physiologic reactions, and in other psychologic capacities, could conceive of erotic capacities (of all things) that were basically uniform throughout a population" (Cited in Irvine 1990, 36). See also Weinrich (2014).

adopted and modified by psychologists and psychiatrists revising the DSM. A central concern of the work of Masters and Johnson was to develop therapies for sexual dysfunctions, including physical problems such as vaginismus (spasms of the pelvic muscles which makes intercourse painful or impossible). Although the sexual response cycle was characterised in strictly physiological terms, Masters thought that sexual dysfunctions were usually due to psychogenic causes.[4]

Critics argued that the human sexual response cycle theorised by Masters and Johnson is less apt for females than it is for males (see Basson 2000; Wood, Koch, and Mansfield 2006; Meana 2010). Their model did not include desire, assuming that desire occurred spontaneously. Though it was dubbed a 'cycle', critics called it 'linear', because it began with arousal and ended with orgasm and resolution. Critics noted that it ignored quality of relationships or other features of a female's social context that can influence sexual experience. More recent theories of female sexual response have attempted to accommodate these considerations. Basson, for example, has argued that female sexual desire is typically responsive (to cues, partner initiation, arousal) rather than spontaneous; that female sexual experience is typically 'circular', in which arousal can lead to desire and satisfaction can generate new desire; and that female sexual desire is modulated by social contexts such as relationship intimacy (see Basson 2000; Meana 2010).

By the late 1970s, the most common form of female sexual dysfunction, the general term for the cluster of diseases of which low female sexual desire is one, was no longer physical problems like vaginismus, but rather involved low sexual desire (Irvine 1990; Kleinplatz 2018). This was the problem that sex therapists were most often seeing in their practice (See Irvine 1990; Everard et al. 2000; and the references in Meana 2010). The disease category for low female sexual desire today is 'female sexual interest/arousal disorder'. To be diagnosed with this disease, four conditions must be met: a female must have at least three of the defining symptoms, the symptoms must persist for at least six months, those symptoms must cause her distress, and the symptoms should not be better explained by other medical conditions or relationship problems or medications. The defining symptoms are an absence of, or reduction in:

- interest in sexual activity,
- sexual thoughts or fantasies,
- initiation of sexual activity and reception of a partner's initiatives,

[4] Their first book was Masters and Johnson (1966). See also Fishman (2007).

- excitement or pleasure during sexual activity in most sexual encounters,
- interest and arousal in response to sexual cues,
- genital or non-genital sensations during sexual activity in most encounters. (DSM-5; see also Brotto 2010).

This alleged disease, along with its predecessor (hypoactive sexual desire disorder), is the focal point for the debate regarding the medicalisation of low female sexual desire.

3. The Mainstream View

The mainstream view regarding the medicalisation of low female sexual desire is that this condition is a genuine disease, and thus it ought to be in the domain of medicine and is an apt target for diagnosis and medical intervention. Sexual functioning is a bodily phenomenon, on the mainstream view, and thus sexual dysfunctions are diseases like other bodily dysfunctions. Low sexual desire can cause various forms of suffering. Since medicine can sometimes help alleviate some forms of suffering, at least when such suffering is caused by a disease, there is a principled reason to think that low female sexual desire should be in the jurisdiction of medicine.

The mainstream view has a wide range of adherents. As we saw in §2, the American Psychiatric Association has codified the condition as a disease in various editions of the DSM. Prominent medical scientists such as Irwin Goldstein and the sisters Laura Berman and Jennifer Berman have for decades promoted low female sexual desire as a disease to be treated with pharmaceuticals. Millions of prescriptions have been written in the United States for off-label testosterone use for low female sexual desire, and two drugs have been approved by the FDA for the condition (flibanserin and bremelanotide), though both have extremely modest beneficial effects and a range of harms (discussed below).[5] In a survey of nearly two thousand professionals attending four medical conferences, 85% believed that hypoactive sexual desire disorder is a genuine medical problem (Bachman 2006).[6] We saw above that a spectrum of scholars have held low female sexual desire to be a disease, from Freud and Kinsey and Masters to Brotto and Basson.

[5] Regarding off-label testosterone prescriptions, see Simes and Snabes (2011)

[6] These were conferences of the American College of Obstetricians and Gynecologists, the Endocrine Society, the North American Menopause Society, and the American Society for Reproductive Medicine.

The primary arguments for the mainstream view are:

The Argument from Suffering
The Appetitive Argument
The Argument from Female Equality
The Argument from Treatment Success

I will assess these arguments in that order.

3.1 The Argument from Suffering

The Argument from Suffering notes the prevalence of females with low sexual desire who experience distress from their condition. This argument is often buttressed by appealing to survey data which suggests that a very large percentage of females experience one or more of the symptoms that constitute the definition of the disease category. One particularly controversial report claimed that 43% of women suffer from some sort of sexual dysfunction (Laumann, Paik, and Rosen 1999; Berman, Berman, and Goldstein 1999; see Moynihan 2003 for criticism of this statistic). Critics claim that this figure is grossly exaggerated. Nevertheless, the most common problem that motivates visits to sex therapists for females is low sexual desire (see Irvine 1990; Kleinplatz 2018). Sometimes the widespread suffering caused by low sexual desire is deployed as a counterargument against the critical view: how insensitive and disrespectful it is to deny treatment to females who suffer.[7] Sometimes this argument is mixed with suggestions of sexism: the scientific study and therapeutic treatment of sex has for long been androcentric, and now we can help males who suffer from erectile dysfunction, while proponents of the critical view are willing to let females suffer in silence.

Though any form of suffering warrants sympathy, as an argument for the mainstream view the Argument from Suffering is question-begging. It assumes as a premise—that low female sexual desire should be in the domain of medicine—the issue which is under dispute. Not all forms of suffering are in the domain of medicine. One need only consider the suffering caused by hunger or climbing high mountains or listening to country music. Even if we grant that low female sexual desire causes suffering, this does not support the mainstream view on medicalisation of low female sexual desire.

[7] Segal (2018) offers a rhetorical analysis of an FDA meeting at which flibanserin was discussed, and she notes that this argument—the suffering caused by an 'unmet medical need'—was one of several offered by promoters of the drug.

Moreover, we will see below that the notion of suffering in this context is contested (§4). Critics hold that the suffering associated with low female sexual desire is typically not an intrinsic harm to the females with the condition, but rather arises as a result of social norms of sexuality or relationship difficulties. To consider an analogy, a homosexual male in present-day Russia might suffer distress from his sexual orientation, not because his sexual orientation is intrinsically harmful (obviously), but because he lives in a society which subordinates and physically harms homosexuals. This rejoinder to the Argument from Suffering is itself inconclusive when deployed against the entire category of low female sexual desire, for reasons we will see in §4, though it is persuasive for some proportion of cases.

3.2 The Appetitive Argument

We saw above that some hold that sexual desire is like an appetite or physiological capacity, and low sexual desire is a result of dysfunction in this capacity. Kinsey, for example, believed that sexual desire is the result of a physiological capacity, akin to the capacity of our pancreas to produce insulin (Kinsey, Pomeroy, and Martin 1948). A low capacity in the latter is a disease (type 1 diabetes), hence a low capacity in the former is also a disease.

A physiological capacity view has been widely adopted by those promoting low female sexual desire as a disease. Some theorise that low female sexual desire is a result of low levels of particular hormones such as testosterone—the Berman sisters are two prominent defenders of the mainstream view who frequently have claimed that low sexual desire in women can be treated with testosterone, and a testosterone patch was being developed for low female sexual desire but was ultimately rejected for consumer use by the FDA (because of concerns about harmful side effects such as heart attacks, breast cancer, and weight gain), though it was approved in Europe. Others theorise that low female sexual desire is a result of an imbalance in neurotransmitters (see for example Croft 2017); this is the basis of the first drug approved for low female sexual desire (flibanserin). After the success of Viagra for erectile dysfunction, its manufacturer began testing it for treating low sexual desire in women. All these attempts to develop pharmaceutical interventions for low female sexual desire assume a physiological capacity view of sexual desire.

One problem with the Appetitive Argument is that it ignores the intentional, psychological, social, and cultural context of sexuality. In §4 I describe some of the substantive ways that this challenge has been articulated, though in §5 I argue that appealing to the causal aetiology of

low female sexual desire offers more modest support to the critical view than its defenders suggest. A further problem with the Appetitive Argument is that thus far no physiological basis for low female sexual desire in general has been discerned.

Nevertheless, it is plausible that for some females with low sexual desire, the cause of their low desire is indeed a result of a dysfunction in a physiological capacity. There are reasons to think that some hormone concentrations can influence sexual desire (in both males and females). We have empirical evidence suggesting that modulating physiological states with pharmaceuticals such as selective serotonin reuptake inhibitors can dampen sexual desire, which itself suggests that sexual desire has a biological basis of one form or another (Bala et al. 2018). Though this consideration might have some initial appeal for a defender of the mainstream view, it is in fact far from conclusive. That is because yet another problem with the Appetitive Argument is that many features of life which are non-medical have a grounding in a physiological capacity. Athletic prowess is a good example. One's running speed is a function not only of training but also of an intrinsic physiological capacity. Alexei's slow running speed might be a function of his unusually low intrinsic physiological capacity for running, but that does not entail that Alexei has a disease.

However, the Appetitive Argument together with the Argument from Suffering are jointly persuasive, for at least some cases of low female sexual desire. It is plausible that some cases of low female sexual desire have a physiological aetiology, and that this causes those people to suffer (though in §4 we see that this latter premise must be understood with care). There is, thus, some reason to think that at least for some cases of low female sexual desire, those cases are genuine diseases.

3.3 The Argument from Female Equality

We saw above that proponents of the mainstream view sometimes frame the medicalisation of female sexual desire as an issue about equality between the sexes. There are grounds for thinking that sex research has been unduly focused on male sexuality. For example, during her research about evolutionary theories of the female orgasm, the philosopher Elizabeth Lloyd traced sociobiologists' footnotes regarding the scientific study of orgasms, and she found that, in the context of theorising about female orgasms, many of the cited sources were in fact based on the study

of males (see Lloyd 2005; Okruhlik 1994).[8] With the success of pharmaceutical treatments for erectile dysfunction beginning in the late 1990s, there was an immediate motivation to develop an equivalent intervention for females. The Argument from Female Equality claims that it is only fair that disorders of female desire receive the same attention as their male equivalents. If male sexual dysfunctions can be medicalised, then so can female sexual dysfunctions. This argument was the basis of the name for the recent industry-funded patient advocacy campaign for the drug flibanserin: Even The Score.[9]

This argument has several damning problems. It assumes that low male sexual desire itself ought to be in the domain of medicine. The argument seems to be: if low male sexual desire has been successfully medicalised, then so too should low female sexual desire. But the critical view on the medicalisation of low female sexual desire applies equally to low male sexual desire—critics have argued that male sexuality has been inappropriately medicalised (Tiefer 1986, 1994; Fishman 2007). Moreover, Bueter and Jukola (2020) convincingly argue that feminism has usually been deployed in criticisms of medicalisation and biological reductionism; therefore to cite concerns about female equality as grounds for upholding the disease status of low female sexual desire, with the ultimate aim of warranting pharmaceutical intervention for the condition, is far-fetched.

Sometimes the Argument from Female Equality is made in the context of discussions about interventions. The argument goes: males have access to effective interventions for their sexual dysfunctioning, and therefore so should females. But what, critics have asked, is the female analogy of intervening on erectile dysfunction? One hypothesis that received some study was: just as pharmaceuticals like Viagra work by increasing blood flow to the penis, perhaps some interventions can increase blood flow to the clitoris. A barrier to this approach, however, is that many empirical studies suggest little correlation between physical signs of arousal in females, such as vaginal blood flow, and subjective feelings of arousal and desire.[10] Similarly, treatment of erectile dysfunction is not in fact an intervention for low male sexual desire, and thus, at least in the context of interventions, the Argument from Female Equality does not bear on whether low female sexual desire should be medicalised.

[8] Taylor (2015) and Angel (2012) note the uneasy and complicated relationship between feminism and the medicalisation of low female sexual desire.
[9] See Segal (2018) for a critical account of various articulations of this argument.
[10] Though such findings have been observed for decades, they have been demonstrated in an elegant series of experiments by Meredith Chivers. See Chivers et al. (2010) for a review.

3.4 The Argument from Treatment Success

Prominent advocates of the mainstream view have claimed that low female sexual desire can be successfully modulated by pharmaceuticals. This, proponents claim, is a reason to think that low female sexual desire should be in the domain of medicine.

Such proclamations of treatment success are laughable in their hyperbolic contradictions of empirical data. Irwin Goldstein, for instance, claimed that when preparing the FDA submission for flibanserin, the worry was not that the drug would be perceived as enhancing female sexual desire too little, but that it would be perceived as enhancing female sexual desire too much—the company did not want to elicit the concern that the drug would be "turning women into nymphomaniacs".[11] The drug in question was rejected by the FDA twice, before it was finally approved during the Even The Score campaign. The basis of the rejections were the tiny observed beneficial effects of the drug, and concerns about its harm profile (one trial testing the safety of this drug to treat low sexual desire in females included only males). Earlier attempts to develop testosterone interventions also floundered upon careful evaluation. The second and thus far last drug approved for low female sexual desire (bremelanotide) has an effect size similar to that of flibanserin. On average, compared with placebo, flibanserin is associated with an increase of about one 'sexually satisfying event' every two months (Jaspers et al. 2016).

4. The Critical View

Critics have argued that low female sexual desire has been inappropriately medicalised. This charge involves a number of related claims: that low female sexual desire is a normal part of life, that low female sexual desire is not caused by medical problems but rather is caused by social, relational, or cultural factors, that the very idea that low female sexual desire is a problem reflects particular social values, that the best way to help low female sexual desire (assuming help is called for) involves non-medical interventions, and that the condition has been constructed as a disease in part because of the financial gains to be had by selling treatments for it.

The critical view has a range of adherents. The New View Campaign, led by psychologist Leonore Tiefer, is among the more visible organisations

[11] In Goldstein's words: "When you're going to the FDA with this kind of drug, there's the sense that you want your effects to be good but not too good (…) there was a lot of discussion about it by the experts in the room, the need to show that you're not turning women into nymphomaniacs. There's a bias, a bias against—a fear of creating the sexually aggressive woman." Cited in Bergner (2014).

defending the critical view, and John Bancroft, the former director of the Kinsey Institute, has also defended the critical view. The journalist Ray Moynihan has published a number of articles and books in which he decries medicalisation practices such as 'disease-mongering' or 'selling sickness', and he has applied such arguments to low female sexual desire. Several academic commentators have aligned themselves with the critical view of medicalising low female sexual desire in scholarly publications (see. e.g. Kaschak and Tiefer 2001; Moynihan 2003; Moynihan and Mintzes 2010; Bancroft 2002; Taylor 2015; Angel 2012; Cacchioni 2015).

The primary arguments for the critical view are:

The Spurious Disease Argument
The Construction of Distress Argument
The Argument from Treatment Failure
The Conflict of Interest Argument
The Harms Argument

I address each in turn, going from subtle to simple.

4.1 The Spurious Disease Argument

Sometimes the debate about the medicalisation of a condition involves the claim that the condition is, or is not, a genuine disease. If a condition is a genuine disease, then, goes this thought, it should be in the domain of medicine; if a condition is not a genuine disease, then there is at least some reason to suppose that the condition should not be in the domain of medicine (though medicine does have in its domain conditions that are not diseases, such as pregnancy). In §3 we saw the Appetitive Argument for the mainstream view. The Spurious Disease Argument for the critical view denies the appetitive model of low sexual desire. Indeed, the charge of medicalisation of low female sexual desire often involves a denial of the capacity view of sexual desire, or at least a denial that the capacity view is a complete explanation for varying strengths of sexual desire. Critics argue that the view of low sexual desire as a deficiency in a physiological capacity is excessively reductionist, and to understand a female's low sexual desire we must take into account that female's broader social context.[12] To properly understand why a female has low sexual desire, one must consider many features of her life, including her general health, levels

[12] See, among many others, Tiefer (1991). Leiblum, for example, claimed that "Inferring that hormones, in general, are the primary motivators of sexual activity in humans is a gross oversimplification" (2002, 65).

of stress, competing interests, and features of her past and present relationships.

Taking this contextual approach further, some feminists such as Catherine MacKinnon (1989) argue that a theory of female sexuality must be located within a broader theory of gender inequality. A proper characterisation of female sexual dysfunction should not begin with the assumption that normal healthy human sexual desire is that of males. Male sexual desire is, obviously, itself influenced by social shaping. Moreover, male and female sexual desire is radically different, claims MacKinnon (which is itself a controversial premise). Females who seem to have dysfunctionally low female sexual desire should instead be seen as resisting a male-centric system and standards of sexuality.[13] Cases of apparent low sexual desire—at least many cases—should be understood, argues MacKinnon and others, as appropriate responses to gender inequality and sexual violence.

A more mundane version of the Spurious Disease Argument was voiced by none other than Lori Brotto, a psychologist who chaired the DSM-5 sexuality committee—the group which developed the disease category 'female sexual interest/arousal disorder'. When interviewed about low female sexual desire, Brotto claimed: "Sometimes I wonder whether it isn't so much about libido as it is about boredom". Brotto was referring to the typical decline in sexual desire that occurs in long-term monogamous relationships.[14]

If the Spurious Disease Argument is meant as a thesis about some token instances of low female sexual desire, then it is convincing, since it is surely plausible that for some females diagnosed with the disease, their condition is better understood as arising from their social context rather than from their intrinsic physiological capacities. However, if the Spurious Disease Argument is meant as a thesis about the disease itself, as a kind, then it is less convincing, since the thesis would deny that any particular instance of low female sexual desire could be a case of disease. That, though, would be committed to claiming that there does not exist a female with low sexual desire for whom their condition is a disease. And that is implausible. To see why, consider what any of the leading philosophical theories of disease must say about a female who, for the sake of argument,

[13] It is a mistake, argues MacKinnon, to see women with low sexual desire "as in need of explanation and adjustment, stigmatized as inhibited and repressed and asexual" (1989, 141)

[14] The Brotto interview is reported in (Bergner 2014). During therapy for women diagnosed with low sexual desire, Brotto noted that "the impact of relationship duration is something that comes up constantly". For this reason, Bergner, who conducted this interview, calls drugs like flibanserin less of an intervention for libido and more of an intervention for monogamy.

suffers genuine distress as a result of her low sexual desire (we will see below that this premise requires nuance).

Normativism about disease holds, roughly, that if a condition is disvalued and if medicine can help, then that condition is a disease. For the Spurious Disease Argument to work as a thesis about the disease itself, assuming normativism, one would have to deny either that the condition is disvalued (but we have granted for the sake of argument that the female in question suffers), or that for all females who experience low sexual desire, medicine cannot help. This latter premise is of course empirical, but it is extremely implausible. Naturalism about disease, on the other hand, holds roughly that if a condition involves a statistical departure from normal functioning, and that dysfunctioning impedes with the ultimate aims of survival and reproduction, then that condition is a disease. For the Spurious Disease Argument to work as a thesis about the disease itself, assuming naturalism, one would have to deny that there exists a female whose sexual desire is much lower than the statistical norm and which impedes her survival or reproduction. This, again, is highly implausible. My favoured account of disease is a hybrid account, which also entails that the Spurious Disease Argument cannot be about the disease itself as a general kind.[15] (It is worth noting that the arguments in this paragraph dodge the question about aetiology altogether—we will return to this in §5.)

To sum: the Spurious Disease Argument may be compelling when understood as thesis about some instances of low female sexual desire, but not when understood about the entire disease category.[16] Of course, among all the females who are diagnosed with a disease of low sexual desire, the proportion for whom the Spurious Disease Argument applies remains an open question. We have seen several reasons to think that for many females who are diagnosed with a disease of low sexual desire, their condition is better understood in social or cultural terms, and so their diagnosis may be inappropriate. Thus, the Spurious Disease Argument provides less warrant to a general thesis of medicalisation of low female sexual desire, and more warrant to what Gabriel and Goldberg (2014) call ‘disease inflation’: the expansion of diagnostic categories and the loosening of diagnostic practices and prescription norms such that more and more people are said to be diseased and are prescribed interventions.

[15] On normativism, see Cooper (2002). On naturalism, see Boorse (1977). On hybridism, see Stegenga (2015).

[16] Some proponents of the critical view are occasionally slippery on this point. Moynihan, for example, claims that while it is surely true that some females have a genuine disease of low sexual desire, the disease category itself is the “freshest, clearest example” of “the corporate sponsored creation of a disease” (2003).

One further nuance is worth mentioning. The above discussion relied on a distinction between condition types and condition tokens: the Spurious Disease Argument fails as a thesis about the condition type (the general category of low female sexual desire), but it might succeed as a thesis about condition tokens (*Sveta's* low sexual desire is not a case of genuine disease, it is a result of an abusive marriage). The underlying premise is that claims of medicalisation should apply to condition tokens rather than condition types, because two people could have the same type of condition in which one of the tokens is constituted by a disease and the other is not. But this would only make sense if by 'condition' one meant 'cluster of symptoms': one cluster of symptoms could be caused by a disease, while another cluster of those same symptoms could be caused by some non-disease state (for example, Maria's sadness and crying and sleeplessness are caused by her depression, while Sofia's sadness and crying and sleeplessness are caused by the recent breakup with her spouse). But if by 'condition' one meant 'whole disease entity, including symptoms and physiological causes of those symptoms', then two tokens of a condition would share all physical features, and thus, arguably, two tokens of the same condition would either both be genuine diseases or both be non-disease conditions. All tokens of type 1 diabetes are cases of genuine disease, while all tokens of appreciating country music are non-disease conditions (though distress-inducing nevertheless). Since the Spurious Disease Argument fails as a thesis about condition types, it can only succeed as a thesis about some condition tokens. But how could it be, following the above line, that some tokens of a condition are genuine diseases while other tokens of the condition are not genuine diseases, if they are tokens of the same condition? One answer which has tempted many defenders of the critical view, and which we have already touched upon, is to distinguish genuine disease tokens from spurious disease tokens according to the aetiology of those tokens. This, finally, brings us to a remaining nuance for Spurious Disease Argument, which I address in §5.

4.2 The Construction of Distress Argument

To be diagnosed with female sexual interest/arousal disorder, the DSM stipulates that a female must suffer distress from her symptoms of low sexual desire. At first glance this seems like a reasonable requirement, since the symptoms alone are not necessarily pathological and it is hard to see what other reason medicine could have to hold that a female with such symptoms is diseased. Indeed, many asexuals have no sexual desire at all and yet do not experience distress as a result, and many would deny that they have a disease. However, the requirement that a female experience distress from her symptoms of low desire in order to be diagnosed raises difficult questions. The Construction of Distress argument holds that the

distress that a female with low sexual desire experiences can be a result of social or cultural features of the female's context, rather than a result of the symptoms themselves (we saw the Construction of Distress Argument foreshadowed as a response to the mainstream position's Argument from Suffering in §3). A female could experience such distress if she felt that she was not satisfying social norms regarding sexual activity or pleasure. Such norms might be generated by manifold social forces, such as peers, advertising, and pornography. Moreover, such norms might be unwarranted or thoroughly pernicious.

The Construction of Distress argument has an additional complexity. Female sexual desire is often deemed low only relative to the strength of their typically male partners. Such distress, in many cases of low female sexual desire, might not be intrinsic, but rather might be relational. That is, such distress can arise not from the female's symptoms directly, but rather from relationship difficulties which arise due to an imbalance of desire with her partner (see, e.g., Irvine 1990).[17]

A curious proviso to the description of female sexual interest/arousal disorder in the DSM-5 notes that there is variability in the prevalence of low sexual desire in different cultures, and cautions:

> A judgement about whether low sexual desire reported by a woman from a certain ethnocultural group meets criteria for female sexual interest/arousal disorder must take into account the fact that different cultures may pathologise some behaviors and not others. (APA 2013, 436)

This appears to be a form of cultural relativism regarding whether a case of low female sexual desire should be deemed a disease or not. One might think that this is muddled, since whether a person has a disease should not depend on culture-specific idiosyncrasies regarding whether that culture pathologizes the condition in question. However, such cultural relativism of disease attribution could be reasonable if it is the case that in some cultures a female with low sexual desire experiences distress while in other cultures a female with low sexual desire experiences no distress, due to differences in the extent to which the cultures pathologises low female sexual desire. But this faces the Construction of Distress argument: the distress that females experience because of the pathologizing tendencies of their culture are, trivially, a result of their culture, and not a result of

[17] Taylor (2015) notes that many of the alleged cases of successful treatment of low female sexual desire described by the Berman sisters involved females who were distressed as a result of partner frustration (Berman, Berman, and Bumiller 2001).

intrinsic harms caused by the condition itself. The DSM is explicitly asserting that the distress caused by low female sexual desire is a cultural construction—a puzzling gesture of support for the critical view from what could be taken as the bible of the mainstream view.

Responding to the Construction of Distress argument, defenders of the mainstream view claim that the argument ignores or trivialises suffering of some females with low sexual desire (see Jackson 2004). Yet, if the source of the distress is indeed a result of the pathologizing tendency of a society, on its face this suggests that diagnosing the condition as a disease and subsequently treating it with biological interventions is misguided. Further, in §5 I argue that the causal aetiology of complex traits such as strength of sexual desire probably involve causes at multiple scales, including both biological and social causes.

4.3 The Argument from Treatment Failure

We have seen that an argument for the mainstream view appeals to claims about the successful treatment of low female sexual desire, and that these claims are empirically implausible. The critical view turns this argument around in the Argument from Treatment Failure, in which the low effectiveness of interventions for low female sexual desire is cited in the context of discussing the condition's medicalisation (see Moynihan 2014). The drugs introduced in the last couple of decades to treat erectile dysfunction are among the most successful pharmacological developments of the last several decades (by various metrics: capacity to modulate the condition, number of prescriptions, number of men taking the drugs, profitability for the manufacturers; but not, obviously, to save lives or mitigate symptoms of mortal diseases). Conversely, only two of many experimental drugs for low female sexual desire have made it through the research and regulatory pipeline, and these drugs have extremely modest beneficial effects for females but significant harms (see below). Drugs to improve low female sexual desire have been failures. One possible explanation for such failures is that the condition is not a genuine disease. The underlying argument is: so far there has been no effective intervention developed for low female sexual desire; if low female sexual desire were a genuine disease, an effective intervention would have, by now, been developed; thus, low female sexual desire is not a genuine disease.

One response to the failure of female desire drugs has been to conclude that female sexuality is complex. Indeed, this appeal to the complexity of female sexual desire formed the basis of criticisms of the development of pharmaceutical interventions for female sexual desire, voiced by academic commentators and feminist advocacy groups, even prior to the empirical

failures of these drugs.[18] No wonder such drugs have been failures, goes this argument: male sexual arousal may be physiologically simple, but female sexual desire is not.

Treatment failure can, of course, be merely transient. Our failure to adequately treat type 1 diabetes until Banting and Best's breakthrough did not entail that type 1 diabetes is not a genuine disease. Thus, the Argument from Treatment Failure is far from conclusive for the critical view. Yet, at the very least the Argument from Treatment Failure is a compelling rejoinder to the mainstream view's Argument from Treatment Success.

Moreover, the failure to modulate female desire with pharmaceuticals is not due to a lack of effort on the part of scientists and companies to find such a drug. The fantastic profits to be gained from a female desire drug have spurred an enormous search. This is a case in which absence of evidence is some evidence of absence.[19] The absence of evidence of effective medical treatments for low female sexual desire is some evidence that there is not going to be an effective medical treatment for low female sexual desire.[20] We have some reason to think, now, that a drug for female sexual desire is not forthcoming. The inability to medically intervene on a condition provides at least some reason for thinking that the condition should not be in the jurisdiction of medicine.

4.4 The Conflict of Interest Argument

Sometimes the charge of medicalisation involves describing tactics used by interested parties in convincing others, especially physicians and potential future patients, that a condition is a disease. These tactics include organising meetings of experts with the aim of defining a disease, sponsoring medical education events to inform physicians about the condition, and performing research which suggests that the condition is under-diagnosed and under-treated (Moynihan 2003; Fishman 2004; Cacchioni 2015). The point of these tactics, of course, is to make money by selling interventions for the condition.[21] Let us call this the Conflict of Interest Argument.

[18] See Bueter and Jukola (2020), who argue that the flibanserin case involved a failure in the uptake of criticism, and thus the requirements of Longino's theory of scientific objectivity were not satisfied.
[19] See Sober (2009) for an articulation of the formal conditions under which absence of evidence is indeed evidence of absence, contrary to standard statistical lore.
[20] Hacking's infamous quip "if you can spray them, then they are real" (1983)—originally perhaps an unintended innuendo but here an unapologetic pun—might be apt here.
[21] As Taylor puts it: "The diagnosis is not about illness or abnormality; it is about making large numbers of people think that they are ill or abnormal so that corporations can profit" (2015).

With respect to the question of medicalisation, an implicit premise of this argument seems to be that such tactics would be unnecessary if the condition were in fact a real disease. However, the same tactics cited in the argument—corporate-funded consensus conferences, medical education, awareness-raising campaigns, patient-advocacy groups—are deployed against genuine diseases, such as breast cancer, HIV, and depression. The Conflict of Interest Argument has some rhetorical sway, but is ultimately inconclusive as a consideration pertinent to medicalisation. That is not to say that conflicts of interest are not an important problem in medicine, in medical research, or in debates about the medical status of some conditions. Holman and Geisler (2018) use the case of flibanserin to show that in FDA consultation meetings, financial conflicts of interest appeared to influence the content of testimony offered by patient advocacy panelists, which in turn probably influenced the FDA decision to approve the drug (see also Segal 2018). Conflicts of interest almost surely had some causal influence on the determination of the putative disease status of low female sexual desire. Yet the same kinds of conflicts of interest are present in many areas of medicine and themselves do not necessarily impugn the medical status of a condition.

4.5 The Harms Argument

The potential harms of the medicalisation of low female sexual desire are numerous. The Harms Argument just says: the potential harms of medicalising low sexual desire are reasons not to medicalise the condition. One class of harms is the various adverse effects of the medical interventions used to treat low female sexual desire. At present this is primarily the drug flibanserin, which has several harmful effects, including fatigue, insomnia, and hypotension.[22] Another kind of harm is the reification of spurious and pernicious norms of sexuality.[23] Reiheld argues that in general medicalisation can have the harm of reification, defined as "a process whereby the ontology of an idea shifts from mere concept to real manifestation" (2010, 77). One way this might occur is via looping effects of human classification, in which those people who are diagnosed with a condition come to see themselves and be seen and treated by others

[22] Taylor argues that "the medical treatment of FSD, as with the medical management of menopause, subjects women to health risks and disciplinary treatments in order to accommodate men and to maintain heterosexual marriages" (2015, 43).

[23] As John Bancroft, former director of the Kinsey Institute, claimed "The danger of portraying sexual difficulties as a dysfunction is that it is likely to encourage doctors to prescribe drugs to change sexual function—when the attention should be paid to other aspects of the woman's life. It's also likely to make women think they have a malfunction when they do not." (Quoted in Moynihan 2003). Wardrope (2015) argues that critiques of medicalisation can involve claiming that medicalisation involves 'hermeneutical injustice'. See also de Vries (2007), Verweij (1999), and Gagné-Julien (2021 this issue of EuJAP).

as fundamentally a kind of person (the kind with that condition), and thereby in various ways they become that kind of person.[24] Medicalising any condition entails a range of financial costs. Finally, attention can be drawn away from the important causes of low female sexual desire.

While these are important consideration, the Harms Argument is far from conclusive, since the medicalisation of all conditions comes with harm. Moreover, as Reiheld (2010) argues, medicalisation can also have benefits that offset or outweigh such harms, such as the demarginalisation of previously marginalised patient groups and destigmatisation of previously stigmatised conditions. Yet, at least in the case of low female sexual desire, and considering the Argument from Treatment Failure, the two arguments suggest that the benefit-harm ratio for medicalising low female sexual desire is poor. I argue in the following section that this pragmatic concern is among the most persuasive, albeit simplest, of the arguments for the critical view.

5. Etiological Models of Low Desire

Thus far we have seen several theories about the aetiology of low female sexual desire. One main family of etiological models is based on physiological capacity for sexual desire, and the other main family of etiological modes is based on social context relevant to sexual desire. Proponents of the mainstream view have tended toward the physiological capacity models, whereas proponents of the critical view have tended toward the social context models.

The physiological family of models states that people's capacity for sexual desire varies, and low sexual desire is simply the result of underlying physiological causes, such as low testosterone levels or an imbalance in neurotransmitters. We saw above that this kind of model was favoured by Kinsey, and it is widely held today by pharmaceutical companies. A version of a social context etiological model for sexuality is the repression model, famously articulated by Freud, which states that people's sexual desires are psychogenic, and can be modulated (mildly or extremely, leading in some cases to paraphilias) by psychological mechanisms. Another version of a social context etiological model is the oppression model, which states that females' sexual desires are modulated by gender inequality, stress, fatigue, and fear of violence. This has been defended by feminists such as MacKinnon. Still another version of a social context etiological model is the boredom model, which states that the strength of

[24] This is Hacking's (1995) "looping effects of human kinds".

sexual desire wanes in particular contexts, especially as a result of relationship duration.

These models are not mutually exclusive, of course—low sexual desire can have multiple aetiologies. However, some of the more prominent defenders of the various models have tended to emphasise one model at the expense of the others. Kinsey, for example, downplayed the importance of social context as an explanation for low sexual desire and emphasised physiological capacity. [25] MacKinnon, conversely, downplayed the importance of physiological capacity and emphasised social context. Yet, all these aetiological models have some initial plausibility.

We saw above that appealing to the aetiology of token instances of low female sexual desire could be a way to distinguish cases of low sexual desire which should be understood as genuine diseases from cases of low sexual desire which should not be understood as genuine diseases. The underlying premise of some appeals to the social context etiological models is that if a female's low sexual desire is due to social or cultural causes, then this female does not have a disease, and thus to diagnose her with a disease amounts to inappropriately medicalising her condition.

As persuasive as this claim may be, this line of argumentation requires care to avoid an ambiguity regarding causation of disease.

Many conditions that people consider to be uncontroversially in the domain of medicine arise from causes that are, ultimately, social or cultural. Car accidents, sporting injuries, drug overdoses, and nuclear reactor meltdowns can all lead to conditions that are medical. In a trivial sense these causes of conditions are all social or cultural artefacts, yet we would not say that the resulting conditions are not genuine diseases. Well-stocked grocery stores and liquor stores and pharmacies are the causes of a wide range of diseases, almost surely more than diseases caused by intrinsic physiological dysfunctions. A person's social context can cause a wide range of genuine diseases.

The distinction between social or cultural causes on the one hand and physiological causes on the other is less sharp than one might suppose. We have some understanding of the pathophysiological mechanisms in which infection with *Mycobacterium tuberculosis* causes symptoms of tuberculosis. But we also have some understanding of the mechanisms in

[25] Kinsey "consistently ignored the ways in which women as a social group may have been taught to avoid or dislike sex and sought biological explanations for their supposedly lower sexual capacity" (Irvine 1990, 40).

which the social context of a prisoner in a crowded jail in Kyrgyzstan causes infection with, and subsequent symptoms of, tuberculosis.[26] It is plausible that for many human conditions such as the strength of one's sexual desire, the etiological causal nexus is extremely complex, and the relevant causes exist at various physical scales, from the chemical to the social, and various temporal scales, from the temporally distal to the temporally proximal.

Perhaps what defenders of the critical view have in mind when they appeal to social or cultural models of aetiology of low female sexual desire is a distinction between proximal causes of a disease and distal causes of a disease. The presence of *Mycobacterium tuberculosis* is a proximal cause of symptoms. But how did the prisoner get infected with this bacterium? To explain this adequately one must cite the distal, social cause: jail overcrowding. This is a small victory for the critical view on medicalisation of low female sexual desire, however, because if our interest is in whether a condition is a genuine disease, then all that matters in our hypothetical case is the proximal cause, namely, the presence of the infectious bacterium. Since infectious diseases are far less controversially held to be genuine diseases, we have an argument that diagnosis by appeal to proximal causes of symptoms, and not distal causes, is not merely sanctioned by medical practice but is in fact normal medical practice. Why should diseases of sexual desire be any different?

To give a concrete example of this in the debate about the medicalisation of low female sexual desire, in an insightful article about the medicalisation of female sexual dysfunction (FSD), Taylor argued that "Much of the problem with FSD seems to arise from lack of education, rather than from something aberrant about the women" (2015, 263). While this is almost certainly true, it is also true for many conditions that are uncontroversial diseases. When Alexei tells Mischa that it is safe to ski on this black diamond ski slope, or that he should take the blue pill rather than the red pill, or that one drives on the left side of the road in Canada, Mischa's resulting dysfunctions arise from a lack of education (both his and Alexei's), rather than anything aberrant about Mischa. And yet those dysfunctions could be genuine diseases.

There is an important analogy with recent debates about depression, and because the pertinent arguments are similar, it is worth considering them. In the DSM-IV, the diagnostic category for depression had a 'bereavement exclusion criterion', such that a person who satisfied the symptomatic

[26] Furman (2017) applies such reasoning to argue that a full understanding of AIDS requires both physiological and social models.

criteria for depression was excluded from a diagnosis of depression if they were bereaving. The thought was that a bereaving person's symptoms of depression are better explained by the fact that they have lost a loved one rather than by the hypothesis that they have a disease (Horwitz and Wakefield 2007). Thus the bereavement exclusion criterion amounted to a consideration of a person's social context when determining if that person has a disease (though the social context that was considered was narrow: there was no 'recently unemployed exclusion criterion' or 'listened to excessive Nick Cave albums exclusion criterion' or 'broke up with girlfriend exclusion criterion'). Some commentators noted that bereavement does not immunise one against depression, and indeed, the loss of a loved one can cause depression—not just apparent symptoms of depression, but depression itself. So when revising the description of the disease category for the next edition of the DSM (DSM-5), the bereavement exclusion criterion was eliminated. Critics who had argued that the bereavement exclusion criterion did not go far enough in considering people's social context were disappointed. However, we have seen that this appeal to social context in determining the status of a condition as a disease is inconclusive.

In the DSM-5, the diagnostic criteria for female sexual interest/arousal disorder also stipulates a diagnostic exclusion criterion, based on social context. It reads as follows:

> If interpersonal or significant contextual factors, such as severe relationship distress, intimate partner violence, or other significant stressors, explain the sexual interest/arousal symptoms, then a diagnosis of female sexual interest/arousal disorder would not be made. (APA 2013, 436)

Here the DSM makes a significant nod to social context aetiological models of low female sexual desire. But just as with depression, the deployment of such exclusion criteria assumes that there is a sharp distinction between social causes and physiological causes of a disease, which, I argued above, is not generally true. Presumably the "significant stressors" referred to in the exclusion criterion could itself cause disease, including low female sexual desire. Perhaps what the APA has in mind is that among cases of low female sexual desire, those cases with clear social-context aetiologies should not be deemed cases of disease, while other cases should be; perhaps the assumption is that the remaining cases have a physiological aetiology. But why assume that the latter have a physiological aetiology? More pressing, why assume that the former do not have a physiological aetiology? We have seen that many conditions can have a social-context aetiology *and* be characterised by underlying

physiological states. Perhaps the APA (and proponents of the critical view) believes that low sexual desire is not one of those kinds of conditions. In any case, this exclusion criterion amounts to holding that one set of possible causes of low sexual desire (biological) should be de-emphasised when another set of possible causes (social) is present.

The most plausible way of making sense of this social-context exclusion criterion for diagnosing low female sexual desire is pragmatic. The exclusion criterion makes sense in the context in which medical interventions can do little good for low female sexual desire in general, while the various factors stipulated as excluding a diagnosis—severe relationship distress, partner violence, or other stressors—can, at least in some cases (one optimistically hopes), be modified, and thus targeting social causes of low sexual desire can do much more good than targeting alleged physiological causes. Flibanserin may not help many females' low sexual desire, but ending an abusive relationship might. Moreover, in addition to the known adverse effects that medications for low female sexual desire have on the body, one might worry about another sort of indirect harm: low desire which is a result of a female's social context (relationship problems or work stress or …) might be a cue to modify this context (modify or end the bad relationship, for example), and medicating away that low desire (assuming that such interventions were in fact effective at increasing sexual desire) could silence this cue, and thus decrease the motive for positive change.

Contrast this with erectile disorder. The DSM description for erectile disorder stipulates a similar exclusion criterion (the symptoms must not be better explained by relationship distress or other stressors). Now imagine Sergei, who is in a distressing relationship and has begun to experience symptoms of erectile dysfunction. His rule-following physician is forbidden from making a diagnosis of erectile disorder, despite the fact that she knows that an effective intervention is available. While it might be prudent for Sergei to reconsider aspects of his relationship, it would be excessively prudish to deny him the effective treatment that is now available, on the grounds that his condition has a social-context etiology.[27] This is not to say that the social-context etiological model is not important for Sergei; the same concern about an unintended mitigation of the motive for positive change applies. My suggestion here is pragmatic: since we have effective and relatively safe interventions for erectile disorder, worrying about whether Sergei has a genuine disease is fussy.

[27] Which might explain why in some jurisdictions, such as the United Kingdom, one can purchase Viagra without a prescription or diagnosis.

This pragmatic consideration—which foregrounds the consequences of deeming a condition a disease and asks whether medicine can effectively intervene on the condition—can inform a general approach to debates about the medicalisation of particular conditions. This approach sidesteps the need to determine whether a condition is a genuine disease according to a general philosophical theory of disease. This pragmatic approach is perhaps what lies at the heart of the critical view of the medicalisation of low female sexual desire, since interventions for low female sexual desire have been essentially failures, and, as the critical view notes, such medicalisation runs the risk of mitigating motivation for changing one's social context. The concern about mitigating one's motive for positive change suggests that there is an ethical dimension to this pragmatic consideration. Both the pragmatic and ethical considerations are about the consequences of *intervening* on low female sexual desire, rather than whether low female sexual desire as a condition is or is not a genuine disease.

6. Conclusion

In my survey of some of the primary arguments for the mainstream view, which holds that low female sexual desire should be under medical jurisdiction, I found most of the arguments on both sides inconclusive. All the arguments for the mainstream view are problematic, which itself lends some support to the critical view, since the status quo has little warrant (§3). However, the Argument from Suffering together with the Appetitive Argument lends some support to the conclusion that at least some cases of low female sexual desire belong in the domain of medicine.

The arguments for the critical view, however, are on somewhat firmer ground (§4). The Construction of Distress Argument, while perhaps not applying to all females with low sexual desire, presumably applies to many. However, both the Spurious Disease Argument and the Construction of Distress Argument involve appeals to social context etiological models of low sexual desire, which, I argued in §5, is less convincing than proponents of the critical view claim.

The most persuasive arguments for the critical view, I argued, involve pragmatic considerations of the harms and benefits of interventions for low female sexual desire. We have good reasons to think that medicine can do little for females with low sexual desire, and we also have good reasons to think that medicalising female sexual desire causes harms, and these considerations, while simpler than the various inconclusive arguments regarding the genuine disease status of low female sexual desire, are

enough to doubt whether low female sexual desire ought to be in the domain of medicine.

Acknowledgments

I am grateful to Anke Bueter, Erin Nash, Saana Jukola, two anonymous referees, and audiences at several conferences and universities for discussion and commentary.

REFERENCES

American Psychiatric Association. 1952. *The Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association.

American Psychiatric Association. 1968. *DSM-II: Diagnostic and Statistical Manual of Mental Disorders*, 2nd edition. Washington, DC: American Psychiatric Association.

American Psychiatric Association. 1980. *DSM-III: Diagnostic and Statistical Manual of Mental Disorders*, 3rd edition. Washington, DC: American Psychiatric Association.

American Psychiatric Association. 1987. *DSM-III-R: Diagnostic and Statistical Manual of Mental Disorders*, 3rd edition, revised. Washington, DC: American Psychiatric Association.

American Psychiatric Association. 1994. *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, 4th edition. Washington, DC: American Psychiatric Association.

American Psychiatric Association. 2000. *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, text revision. Washington, DC: American Psychiatric Association.

American Psychiatric Association. 2013. *DSM-5: Diagnostic and Statistical Manual of Mental Disorders*, 5th edition. Washington, DC: American Psychiatric Association.

Angel, Katherine. 2010. "The History of 'Female Sexual Dysfunction' as a Mental Disorder in the 20th Century." *Current Opinion in Psychiatry* 23 (6): 536–41. https://doi.org/10.1097/YCO.0b013e32833db7a1.

———. 2012. "Contested Psychiatric Ontology and Feminist Critique: 'Female Sexual Dysfunction' and the *Diagnostic and Statistical Manual*." *History of the Human Sciences* 25 (4): 3–24. https://doi.org/10.1177/0952695112456949.

Bachmann, Gloria. 2006. "Female Sexuality and Sexual Dysfunction: Are We Stuck on the Learning Curve?" *The Journal of Sexual*

Medicine 3 (4): 639–45. https://doi.org/10.1111/j.1743-6109.2006.00265.x.

Bancroft, John. 2002. "The Medicalization of Female Sexual Dysfunction: The Need for Caution." *Archives of Sexual Behavior* 31 (5): 451–55. https://doi.org/10.1023/A:1019800426980.

Basson, Rosemary. 2000. "The Female Sexual Response: A Different Model." *Journal of Sex & Marital Therapy* 26 (1): 51–65. https://doi.org/10.1080/009262300278641.

Bergner, Daniel. 2014. *What Do Women Want?: Adventures in the Science of Female Desire*. CCCO: an imprint of CollinsHarpers Publishers

Berman, Jennifer, Laura Berman, and Elisabeth Bumiller. 2001. *For Women Only: A Revolutionary Guide to Reclaiming Your Sex Life*. New York: Henry Holt and Co.

Berman, Jennifer R, Laura Berman, and Irwin Goldstein. 1999. "Female Sexual Dysfunction: Incidence, Pathophysiology, Evaluation, and Treatment Options." *Urology* 54 (3): 385–91. https://doi.org/10.1016/S0090-4295(99)00230-7.

Boorse, Christopher. 1977. "Health as a Theoretical Concept." *Philosophy of Science* 44 (4): 542–73. https://doi.org/10.1086/288768.

Brotto, Lori A. 2010. "The DSM Diagnostic Criteria for Hypoactive Sexual Desire Disorder in Women." *Archives of Sexual Behavior* 39 (2): 221–39. https://doi.org/10.1007/s10508-009-9543-1.

Bueter, Anke, and Saana Jukola. 2020. "Sex, Drugs, and How to Deal with Criticism: The Case of Flibanserin." In *Uncertainty in Pharmacology*, edited by Adam LaCaze and Barbara Osimani, 338:451–70. Boston Studies in the Philosophy and History of Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-29179-2_20.

Cacchioni, Thea. 2015. *Big Pharma, Women, and the Labour of Love*. University of Toronto Press. https://doi.org/10.3138/9781442694101.

Chivers, Meredith L., Michael C. Seto, Martin L. Lalumière, Ellen Laan, and Teresa Grimbos. 2010. "Agreement of Self-Reported and Genital Measures of Sexual Arousal in Men and Women: A Meta-Analysis." *Archives of Sexual Behavior* 39 (1): 5–56. https://doi.org/10.1007/s10508-009-9556-9.

Cooper, Rachel. 2002. "Disease." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 33 (2): 263–82. https://doi.org/10.1016/S0039-3681(02)00018-3.

Croft, Harry A. 2017. "Understanding the Role of Serotonin in Female Hypoactive Sexual Desire Disorder and Treatment Options." *The Journal of Sexual Medicine* 14 (12): 1575–84.

https://doi.org/10.1016/j.jsxm.2017.10.068.

Fishman, Jennifer. 2007. *Making Viagra: From Impotence to Erectile Dysfunction*. New York: Medicating Modern America: Prescription Drugs in History. A. Tone and E. S. Watkins. New York University Press.

Fishman, Jennifer R. 2004. "Manufacturing Desire: The Commodification of Female Sexual Dysfunction." *Social Studies of Science* 34 (2): 187–218. https://doi.org/10.1177/0306312704043028.

Foucault, Michel. 1979. *The History of Sexuality, Volume 1: An Introduction.* Allen Lane.

Freud, Sigmund. 1905. *Three Essays on the Theory of Sexuality*. Standard Edition of the Complete Psychological Works of Sigmund Freud. Trans. James Strachey. Basic Books. https://books.google.hr/books?id=AUvjxQEACAAJ.

Furman, Katherine. 2020. "Mono-Causal and Multi-Causal Theories of Disease: How to Think Virally and Socially about the Aetiology of AIDS." *Journal of Medical Humanities* 41 (2): 107–21. https://doi.org/10.1007/s10912-017-9441-9.

Gabriel, Joseph M., and Daniel S. Goldberg. 2014. "Big Pharma and the Problem of Disease Inflation." *International Journal of Health Services* 44 (2): 307–22. https://doi.org/10.2190/HS.44.2.h.

Gagné-Julien, Anne-Marie. 2021. "Wrongful Medicalization and Epistemic Injustice in Psychiatry: The Case of Premenstrual Dysphoric Disorder." *European Journal of Analytic Philosophy* 17. https://doi.org/10.31820/ejap.17.3.3.

Hacking, Ian. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge, UK: Cambridge University Press.

———. 1995. "The Looping Effects of Human Kinds." In *Causal Cognition*, edited by Dan Sperber, David Premack, and Ann James Premack, 351–83. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524021.003.0012.

Holman, Bennett, and Sally Geislar. 2018. "Sex Drugs and Corporate Ventriloquism: How to Evaluate Science Policies Intended to Manage Industry-Funded Bias." *Philosophy of Science* 85 (5): 869–81. https://doi.org/10.1086/699713.

Horwitz, Allan V., and Jerome C. Wakefield. 2007. *The Loss of Sadness: How Psychiatry Transformed Normal Sorrow into Depressive Disorder*. Oxford: Oxford University Press.

Irvine, Janice M. 1990. *Disorders of Desire: Sexuality and Gender in Modern American Sexology*. Rev. and Expanded ed. Philadelphia, Pa: Temple University Press.

Jaspers, Loes, Frederik Feys, Wichor M. Bramer, Oscar H. Franco, Peter Leusink, and Ellen T. M. Laan. 2016. "Efficacy and Safety of

Flibanserin for the Treatment of Hypoactive Sexual Desire Disorder in Women: A Systematic Review and Meta-Analysis." *JAMA Internal Medicine* 176 (4): 453. https://doi.org/10.1001/jamainternmed.2015.8565.

Kaschak, Ellyn, and Leonore Tiefer. 2001. *A New View of Women's Sexual Problems*. New York; London: The Haworth Press. http://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9781317788140.

Kinsey, Alfred Charles, Wardell Baxter Pomeroy, and Clyde Eugene Martin. 1948. *Sexual Behavior in the Human Male*. 15. printing. Philadelphia: Saunders Company.

Kleinplatz, Peggy J. 2018. "History of the Treatment of Female Sexual Dysfunction(s)." *Annual Review of Clinical Psychology* 14 (1): 29–54. https://doi.org/10.1146/annurev-clinpsy-050817-084802.

Lloyd, Elisabeth Anne. 2005. *The Case of the Female Orgasm: Bias in the Science of Evolution*. Cambridge, Massachusetts: Harvard University Press.

MacKinnon, Catharine A. 1991. *Toward a Feminist Theory of the State*. Cambridge, Massachusetts: Harvard University Press.

Masters, William, and Virginia Johnson. 1966. *Human Sexual Response*. Boston: Little, Brown and Company.

Meana, Marta. 2010. "Elucidating Women's (Hetero)Sexual Desire: Definitional Challenges and Content Expansion." *Journal of Sex Research* 47 (2–3): 104–22. https://doi.org/10.1080/00224490903402546.

Moynihan, R. 2003. "The Making of a Disease: Female Sexual Dysfunction." *BMJ* 326 (7379): 45–47. https://doi.org/10.1136/bmj.326.7379.45.

Moynihan, Ray. 2014. "Evening the Score on Sex Drugs: Feminist Movement or Marketing Masquerade?" *BMJ* 349 (oct17 8): g6246–g6246. https://doi.org/10.1136/bmj.g6246.

Moynihan, Ray, and Barbara Mintzes. 2010. *Sex, Lies, and Pharmaceuticals: How Drug Companies Plan to Profit from Female Sexual Dysfunction*. Greystone Books.

Okruhlik, Kathleen. 1994. "Gender and the Biological Sciences." *Canadian Journal of Philosophy Supplementary Volume* 20: 21–42. https://doi.org/10.1080/00455091.1994.10717393.

Reiheld, Alison. 2010. "Patient Complains of …: How Medicalization Mediates Power and Justice." *IJFAB: International Journal of Feminist Approaches to Bioethics* 3 (1): 72–98. https://doi.org/10.3138/ijfab.3.1.72.

Segal, Judy Z. 2018. "Sex, Drugs, and Rhetoric: The Case of Flibanserin for 'Female Sexual Dysfunction.'" *Social Studies of Science* 48 (4): 459–82. https://doi.org/10.1177/0306312718778802.

Simes, Stephen M., and Michael C. Snabes. 2011. "Reaction to the Recent Publication by Rosemary Basson Entitled 'Testosterone Therapy for Reduced Libido in Women.'" *Therapeutic Advances in Endocrinology and Metabolism* 2 (2): 95–96. https://doi.org/10.1177/2042018811404034.

Stegenga, Jacob. 2015. "Effectiveness of Medical Interventions." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 54: 34–44. https://doi.org/10.1016/j.shpsc.2015.06.005.

Taylor. 2015. "Female Sexual Dysfunction, Feminist Sexology, and the Psychiatry of the Normal." *Feminist Studies* 41 (2): 259. https://doi.org/10.15767/feministstudies.41.2.259.

Tiefer, Leonore. 1986. "In Pursuit of the Perfect Penis: The Medicalization of Male Sexuality." *American Behavioral Scientist* 29 (5): 579–99. https://doi.org/10.1177/000276486029005006.

———. 1994. "The Medicalization of Impotence: Normalizing Phallocentrism." *Gender & Society* 8 (3): 363–77. https://doi.org/10.1177/089124394008003005.

Verweij, Marcel. 1999. "Medicalization as a Moral Problem for Preventive Medicine." *Bioethics* 13 (2): 89–113. https://doi.org/10.1111/1467-8519.00135.

Vries, Jantina de. 2007. "The Obesity Epidemic: Medical and Ethical Considerations." *Science and Engineering Ethics* 13 (1): 55–67. https://doi.org/10.1007/s11948-007-9002-0.

Wardrope, Alistair. 2015. "Medicalization and Epistemic Injustice." *Medicine, Health Care and Philosophy* 18 (3): 341–52. https://doi.org/10.1007/s11019-014-9608-3.

Weinrich, James D. 2014. "Notes on the Kinsey Scale." *Journal of Bisexuality* 14 (3–4): 333–40. https://doi.org/10.1080/15299716.2014.951139.

European Journal of Analytic Philosophy EuJAP | Vol. 17 | No. 2 | 2021
https://doi.org/10.31820/ejap.17.3.5

WHEN A HYBRID ACCOUNT OF DISORDER IS NOT ENOUGH: THE CASE OF GENDER DYSPHORIA

Kathleen Murphy-Hollies[1]

[1] University of Birmingham

Original scientific article – Received: 31/3/2021 Accepted: 11/10/2021

ABSTRACT

In this paper I discuss Wakefield's account of mental disorder as applied to the case of gender dysphoria (GD). I argue that despite being a hybrid account which brings together a naturalistic and normative element in order to avoid pathologising normal or expectable states, the theory alone is still not extensive enough to answer the question of whether GD should be classed as a disorder. I suggest that the hybrid account falls short in adequately investigating how the harm and dysfunction in cases of GD relate to each other, and secondly that the question of why some dysfunction is disvalued and experienced as harmful requires further consideration. This masks further analysis of patients' distress and results in an unhelpful overlap of two types of clinical patients within a diagnosis of GD; those with gender-role dysphoria and those with sex dysphoria. These two conditions can be associated with different harms and dysfunctions but Wakefield's hybrid account does not have the tools to recognise this. This misunderstanding of the sources of dysfunction and harm in those diagnosed with GD risks ineffective treatment for patients and reinforcing the very same prejudiced norms which were conducive to the state being experienced as harmful in the first place. The theory needs to engage, to a surprising and so far unacknowledged extent, with sociological concepts such as the categorisation and stratification of groups in society and the mechanism of systemic oppression, in order to answer the question of whether GD should be classed as a mental disorder. Only then can it successfully avoid pathologising normal or expectable states, as has been seen in past 'illnesses' such as homosexuality and 'drapetomania'.

Keywords: *mental disorder; Wakefield; hybrid; gender dysphoria; DSM*

Correspondence: KLM276@student.bham.ac.uk

1. Introduction

Gender dysphoria (GD) is commonly seen to underlie trans-identities in transgender people. Despite intense debate regarding whether the condition should be seen as a disorder and included in the DSM, GD was included in the DSM-5. I will assume for present purposes that the DSM aims to catalogue and include only disorders, while allowing that medicine as a wider discipline may reasonably treat conditions which are not strictly disorders and may not be in the DSM. Viewing GD as a mental disorder and including it in the DSM-5 on this basis was opposed by some who argued that the condition is not a disorder and is instead just socially disvalued (Giordano 2013, 55), and that its inclusion therefore reinforced the stigmatization of gender-variant individuals, forcing them to 'meet' a clinical threshold instead of recognizing that perfectly happy and well-functioning gender variant and transgender individuals exist (Lev 2006, 48, 56). Furthermore, others argued that the classification was inherently sexist and misogynistic, pathologising those who exhibit atypical gender behaviour and pushing 'patients' into conforming rather than self-acceptance (Langer and Martin 2004, 14-15). This would be a contemporary echo of the pathologisation of homosexual people when homosexuality was included in the DSM-II and DSM-III.

I will explore whether GD should be classed as a disorder and therefore included in the DSM-5, and specifically whether using Wakefield's hybrid account of disorder helps clarify this issue. Or in other words, whether Wakefield's hybrid account helps us to delineate between a socially disvalued state, and a disorder which ought to be included in the DSM. Wakefield's hybrid account is a hugely influential account of mental disorder (see Faucher & Forest 2021), which is still discussed in relation to and applied to, for example, cases of delusions (Miyazono 2015; Lancelotta and Bortolotti 2020), misbelief (McKay & Dennett 2009), psychopathy (Jurjako 2019), and autism spectrum disorder (Wakefield, Wasserman, and Conrad 2020).

Importantly, Wakefield claims that his hybrid account avoids psychiatry's historical problem of pathologising disvalued natural states (such as homosexuality) by tying the harm that an individual experiences to a dysfunction, the identification of which requires no value judgements. He says that "The harmful dysfunction view allows us to reject these diagnoses on scientific grounds, namely, that the beliefs about natural functioning that underlie them (…) are false" (Wakefield 1992, 386). It is this claim, that the incorporation of these two elements successfully picks

out socially disvalued states from those which are truly disordered, that I challenge.

The danger of pathologising natural states just because they are socially disvalued is more widely recognised in the context of normative accounts of disorder, such as Nordenfelt's (2007). In the case of GD, rates of GD may fluctuate depending on how accepting the surrounding environment of the individual is and treatment could force the patient into conforming to non-ideal cultural standards. Naturalist approaches to defining mental disorder such as Boorse's (1975, 57) use scientific markers of disorder such as the loss of natural functions which are detrimental to survival and reproduction. However, I show that the case of GD and its relation to the sociology of gender demonstrates how, fundamentally, sociology frames what can be coherently identified as a dysfunction at all. Therefore, another reason I use Wakefield's hybrid account is that if GD represents a problem for the hybrid account, similar problems will apply to these other accounts of disorder.

I argue that the complex case of GD demonstrates the extent to which a successful account of what constitutes a mental disorder will have to engage with sociological discourses, such as those regarding the stratification of groups in society and how systematic oppression occurs, in order to end psychiatry's troubled history of pathologising normal and healthy states (for discussions of other cases of medicalization, see Gagné-Julien 2021 and Stegenga 2021 in this issue of EuJAP). Even Wakefield's hybrid account does not do this, and so despite tying a normative harm to a naturalistic dysfunction in order to avoid pathologising socially disvalued states the theory is still not comprehensive enough to do so successfully. When it comes to gender, what kind of understanding of gender we adopt determines whether the classification for GD accurately identifies a disorder, or whether it merely reflects and reinforces harmful social norms and expectations.[1] Wakefield's claim that the hybrid account avoids pathologising natural states is shown to be false, as further sociological engagement is required. Whether this element could be incorporated into some neo-hybrid account of disorder or an entirely new approach is needed, I do not specify.

[1] There is discussion that the use of the term "disability" in the DSM-5 may implicitly draw this distinction between disorder and social disability (Cooper 2018). In the case of GD, it may be that the condition should be understood as primarily a disability, but this is not made clear in the DSM-5 and the potentially harmful consequences I discuss, particularly regarding treatment, could still follow.

2. Wakefield's Harm and Dysfunction Analysis

Wakefield's hybrid account brings together a factual value-free component and culturally determined value-laden component, in an attempt to capture the best parts of each in analysing the concept of mental disorder. The first component is the requirement of a dysfunction in a (mental) mechanism, whereby it is no longer carrying out its natural function (Wakefield 1992, 382). According to Wakefield, these natural functions can be identified by reference to earlier evolutionary pressures which would have caused these mechanisms to exist and function in the way that they do. This would have been because they somehow aided the survival and/or reproduction of humans in the past. This process of identifying a dysfunction can therefore be difficult because it will require theorizing about the evolutionarily adaptive nature of various mechanisms, but should be a "purely factual scientific" matter (Wakefield 1992, 383). This may involve measuring the output of a mechanism and comparing it with the optimal level of functioning of that mechanism in order to determine whether it is fulfilling its natural function.

Whether it is in fact possible to identify dysfunction in such a value-free way is a matter of controversy, given that many mechanisms present in humans today perform useful functions which they were not originally 'designed' by evolution to perform (Lilienfeld and Marino 1995, 412) or are 'spandrels'—by-products from the development of other useful mechanisms (Murphy and Woolfolk 2000, 243). But for present purposes, I aim to show that the move of positing a value-free dysfunction as the source of harm in some condition will be insufficient in delineating disorder from disvalued state, for reasons that do not solely relate to the presence of value judgements.

Due to the fact that many of us will have some degree of dysfunction in various psychological processes which are in fact harmless and which we may not even be aware of, Wakefield's harm requirement must also be met for a condition to be classed as a mental disorder. To ascertain whether a dysfunction is harmful, we must apply cultural values of harm and societal expectations of what is a good quality of life (Wakefield 1992, 383-384). Essentially, only mental dysfunctions that stop someone from living healthily and comfortably, constitute mental disorders.

Wakefield (1992, 386) argues that these two components together avoid pathologising natural states. In the past, pathologising natural states has caused great harm to individuals, as is seen in the case of homosexuality. These individuals may feel pressured to suppress manifestations of the

'condition' and struggle deeply with accepting themselves, significantly reducing their well-being. By specifying that the harm and distress experienced with a condition must be caused by the dysfunction, the presence of which is identified without any value judgements, Wakefield claims to avoid the pathologisation of natural states such as homosexuality just because those conditions are disvalued in society. The distress often experienced by homosexual individuals is caused exclusively by prejudice and hostility from the surrounding society, not from any dysfunction. This demonstrates that the dysfunction must be solely 'in the individual', such that if the truly disordered individual were removed from the society to live alone, harm and distress would still be experienced by them because it is tied to the dysfunction within themselves. The distress, therefore, "cannot be due to social deviance, disapproval by others, or conflict with society or others" (Wakefield and First, 2003, 34).

3. Applying Harm and Dysfunction to Gender Dysphoria

Before moving on to Gender dysphoria in the DSM-5, I will briefly discuss Gender Identity disorder (GID) in the DSM-IV-TR. It is defined as a condition in which an individual experiences a gender identity which conflicts with their external sexual characteristics and associated gender role, and therefore suffers gender dysphoria. It involves a "strong and persistent cross-gender identification (not merely a desire for any perceived cultural advantages of being the other sex)." (DSM-IV-TR, American Psychiatric Association, APA, 2000, 581). For children to be diagnosed with the disorder, they must meet 4 of the following criteria:

1. Repeatedly stated desire to be, or insistence that he or she is, the other sex.
2. In boys, preference for cross-dressing or simulating female attire; in girls, insistence on wearing only stereotypical masculine clothing.
3. Strong and persistent preferences for cross-sex roles in make-believe play or persistent fantasies of being the other sex.
4. Intense desire to participate in the stereotypical games and pastimes of the other sex.
5. Strong preference for playmates of the other sex.

The DSM also describes a "Persistent discomfort with his or her sex or sense of inappropriateness in the gender role of that sex", which may manifest in boys and girls asserting that their genitalia are disgusting and that they would prefer not to have them. Similarly, girls may reject the

reality of upcoming pubertal changes such as breast growth and menstruation. Finally, the condition must not be concurrent with a physical intersex condition and must cause "clinically significant distress or impairment in social, occupational, or other important areas of functioning" (DSM-IV-TR, APA, 2000, 581).

Wakefield has claimed that by specifying that the condition must not merely be a desire for the perceived cultural advantages of being the other sex, GID is included in the DSM-IV-TR in such a way that it successfully takes cultural context into account and therefore avoids a 'false positive', a diagnosis of disorder where there is none (Wakefield and First 2012, 133). He says that we don't necessarily need to know the intricate details of a mechanism at work in order to figure out its natural function (Wakefield 1992, 382), and that GID is one such disorder which "clearly corresponds to a type of inferred designed mechanism that has gone wrong" (Wakefield and First 2003, 36), even if we do not know the intricacies the mechanism of gender development. So, it appears that Wakefield accepts that there is dysfunction in the case of GID.

In terms of harm and impairment, the 2015 US transgender survey found that 39% of transgender individuals reported serious psychological distress, 40% had attempted suicide in their lifetime, 30% had experienced homelessness, 29% were living in poverty and a higher proportion of respondents were unemployed than in the general population (James et al 2016, 10, 13). It is also well-documented that dysphoric feelings of "being wrongly embodied" are extremely distressing, often to the extent that they motivate expensive and risky cosmetic procedures and even self-surgery (Lawrence 2011, 652). These findings suggest that those who are dysphoric with regards to their gender suffer impaired functioning. Given the prevalence of discrimination towards gender variant and transgender individuals, it could be questioned whether these effects are caused by a dysfunction alone. But on a more personal and direct level, those with GID report constant grief and distress associated with having to pretend to be and be perceived as someone they're not, and describe relief when they finally feel able to express themselves with their preferred clothes/pastimes etc. (Giordano 2013, 144). So, overall, it would seem that GID causes harm according to the standards of our culture, and so would count as mental disorder on Wakefield's account.

I maintain that the classification of GD in the DSM-5 is similar enough that these claims to harm and dysfunction, and Wakefield's comments about GID, would also apply to GD. In the DSM-5, GD is described as "a marked incongruence between the gender they have been assigned to

(usually at birth, referred to as natal gender) and their experienced/expressed gender" and there must be "evidence of distress about this incongruence" (DSM-5, APA 2013, 453). The specific requirements for a diagnosis are different for children and for adolescents/adults, but for both they must last at least 6 months. For children, a diagnosis of GD requires six of the following with "associated significant distress or impairment in function":

1. A strong desire to be of the other gender or an insistence that one is the other gender.
2. A strong preference for wearing clothes typical of the opposite gender.
3. A strong preference for cross-gender roles in make-believe play or fantasy play.
4. A strong preference for the toys, games or activities stereotypically used or engaged in by the other gender.
5. A strong preference for playmates of the other gender.
6. A strong rejection of toys, games and activities typical of one's assigned gender.
7. A strong dislike of one's sexual anatomy.
8. A strong desire for the physical sex characteristics that match one's experienced gender.

For adolescents, they require two of the following:

1. A marked incongruence between one's experienced/expressed gender and primary and/or secondary sex characteristics.
2. A strong desire to be rid of one's primary and/or secondary sex characteristics.
3. A strong desire for the primary and/or secondary sex characteristics of the other gender.
4. A strong desire to be of the other gender.
5. A strong desire to be treated as the other gender.
6. A strong conviction that one has the typical feelings and reactions of the other gender. (DSM-5, APA 2013, 452).

I take this account of GD in DSM-5 to be similar enough to the account of GID in DSM-IV-TR to assume that Wakefield's conclusion that GD is a disorder would still apply. Both entries contain diagnostic criteria describing patients insisting that they are the other gender, preferring toys and pastimes associated with the opposite gender, experiencing discomfort with their physical bodies, as well as general distress and impairment. Although the description for GD does not include so explicitly the

requirement that the condition is not just a desire for any perceived cultural advantages of being the other sex, as the criteria for GID does, the updated definition of mental disorder in the DSM-5 states that

> Socially deviant behavior (e.g., political, religious, or sexual) and conflicts that are primarily between the individual and society are not mental disorders unless the deviance or conflict results from a dysfunction in the individual. (DSM-5, APA 2013, 20)

The inclusion of this statement could be seen to express an intention for states which are *solely* reactions to a prejudiced society to not be mistakenly classed as disorders, as would have been the case described by DSM-IV-TR if someone were identifying as another gender for the perceived cultural benefits. Finally, both criteria comprise a mix of two types of symptoms, those which relate to patients having strong preferences for things which are commonly associated with the opposite gender, and those which relate to patients experiencing intense discomfort with their physical, sexed body.

4. Inadequacies

4.1 Dysfunction

I propose that the link between a dysfunction and *all* the symptoms we see in the diagnostic criteria for GD is hard to see and is not accurately identified by applying a hybrid account of disorder. Wakefield refers to a dysfunction when he says that GID "clearly corresponds to a type of inferred designed mechanism that has gone wrong" (Wakefield and First 2003, 36), but does this dysfunction explain both having a preference for certain clothes and pastimes and an intense discomfort with parts of your body?

Some symptoms relate to being profoundly uncomfortable with parts of one's anatomy, and in particular one's primary and secondary sex characteristics. I refer to this discomfort as sex dysphoria. Other symptoms relate to preferences for and rejections of certain clothes, toys, pastimes, even certain feelings and reactions which have close associations with the opposite gender. I refer to this discomfort as gender-role dysphoria. It is important to note that according to the GD criteria, a child can be diagnosed with GD without any symptoms of discomfort with their biological sex, and adolescents can receive a diagnosis of GD whether their

symptoms are solely related to gender roles or solely related to their physical bodies.

So, I suggest that there are two distinct clinical groups with different symptoms and experiences which are muddled together in the disparate diagnostic criteria for GD. It is difficult to draw conclusions from clinical data on the co-occurrence of these distinct phenomena as studies vary in exactly how they define and measure each, but Bentler, Rekers and Rosen found a correlation of 0.7 between "behaviour disturbance" (similar to what I would consider 'gender-role dysphoria') and "identity disturbance" (similar to what I would call 'sex dysphoria'), "thus verifying that behaviour and identity disturbance were highly related but not synonymous phenomena" (1979, 277). Bartlett et al. (2000, 758) consider the possibility that children who have symptoms akin to sex dysphoria may then be expected by others to develop gender-role dysphoria. Another related observation is that many gender-variant and transgender individuals now increasingly present with a vast array of different desires and identities, seeking different surgeries, treatment, or no intervention at all (Lev 2006, 46).

When considering what kind of mental mechanism might have a dysfunction which gives rise to GD, it could be said to be easier to imagine what kind of dysfunction might underlie sex dysphoria. This is partly due to the existence of similar mental disorders which also appear to manifest malfunction in the mental conceptualization of bodily constitution. In these conditions, we encounter an "inferred designed mechanism" (Wakefield and First 2003, 36) for the conceptualization of the boundaries of one's own body. The natural function of this mechanism, we can quite confidently theorize, is significantly evolutionarily adaptive. Lawrence (2006) suggests that a discomfort with one's sex characteristics is a dysfunction within the individual which may be akin to other mental disorders such as Body Dysmorphic Disorder (BDD) or Body Integrity Disorder (BID), and that it is in the presence of a sexist society that those with sex dysphoria end up, as a response to that sex dysphoria, forming new corresponding 'gender identities'. Given this, and the fact that sex dysphoria usually precedes gender-role discomfort in these patients by as much as many years, she argues that symptoms which relate to discomfort with gender roles (i.e., what I call gender-role dysphoria) should be viewed as an epiphenomenon to sex dysphoria, and not an underlying dysfunction or mental disorder itself (see Lawrence 2011, 653).

I also suggest that we are not so inclined to say that those with only sex dysphoria would no longer suffer if they were taken away from a

prejudiced society, and that therefore this appears to be a harmful dysfunction which is based 'in the individual' rather than being a conflict between an individual and society. We have seen how intensely uncomfortable individuals with sex dysphoria can feel towards their sex characteristics and the lengths some go to in an attempt to relieve that discomfort. But it may be a different story when it comes to imagining those with only gender-role dysphoria being removed from a society with any recognizable gender roles. Should we think that a kind of bodily-conception dysfunction also explains gender role-dysphoria, and therefore all of GD? I believe that an answer to this question necessarily involves looking at how the notion of 'gender' should be understood.

4.2 Two Understandings of Gender

A full and comprehensive exploration of all the available attempts in the literature to give an account of what 'gender' is would be beyond the scope of this paper, but I suggest that a few differing key aspects would have significant repercussions on our understanding of GD. Here I present two basic conceptions of 'gender' with some key differences which relate to the ontological status of gender, the sex and gender distinction, and whether gender is wholly *harmful* gender roles.

A first account of gender which I'll consider, the 'traditional account' of gender, understands it to be an external set of cultural roles, traits and expectations (from here on, 'gender roles') which are projected and imposed onto people in society through socialisation, with an individual's sex determining which roles and expectations will be imposed. This notion of gender is associated with second-wave feminism and reflected in the feminist slogan that "gender is the social significance of sex", where sex is a basic biological category. De Beauvoir's well-known statement that "One is not born, but rather becomes a woman" (1949, found in 1997) is widely regarded as the birth of the distinction between sex and gender (Ásta 2018, 42), despite the fact the de Beauvoir is now generally interpreted not to have endorsed an account which juxtapositions sex and gender as such separate and different categories (see Ásta 2018; Moi 1999; though also Gatens 2003 for a closer examination of the status of 'biological sex' in de Beauvoir's work). Nevertheless, this traditional account is committed to a distinction between gender roles and the sexed body, such that gender roles are hung on the "coat-rack" (Nicholson 1994, 81) of one's biological sex; the gender roles imposed constitute your gender and it therefore is not self-generated.

Importantly, these gender roles are more liberating and preferential for

men, while oppressive and harmful for women. The gender roles reinforce women's subordination (Millett 1971, 26) and so women are oppressed through having to 'be' women, by having to abide by these gender roles. Therefore, we should work towards a genderless (though not sexless) world (Rubin 1975). Given that these roles are, however, essentially cultural, not only can they in principle be changed or eradicated, but the category of 'woman' is more likely to be defined on the basis of a hierarchical position which women hold, rather than anything else. In Haslanger's (2000) ameliorative enquiry, for example, women are defined as those who occupy a subordinate social position, as this definition best suits political feminist aims.

A second account of gender which I'll consider, an 'identity-based' view of gender, differs from the previous in some key respects. This account understands someone's gender to be a part of their identity, in some form, which in turn tells them which gender roles are appropriate for them. It appears to be internally generated and then has an important link to being expressed with certain perceived gendered hobbies, clothes, feelings etc. So, in reverse to the traditional account, on this account a sense of gender precedes the gender roles. We see this kind of understanding of gender in play quite explicitly in political steps towards prioritising the value of self-identification of gender in gender-variant individuals (Fairbairn, Pyper, Gheera and Loft, 2020).

This shift in understanding gender is reflected in Butler's work post-Beauvoir. Firstly, she reevaluates the ontological statuses of sex and gender. In the traditional account, the value-free scientific matter of one's sex determines one's gender by determining which culturally sanctioned gender roles are imposed. However, on Butler's (1990) account, these cultural ideas about gender roles actually form and regulate the categories of sex. She states that what gives sex categories meaning and makes them intelligible to us are shared cultural ideas about gender, such that "Gender ought not to be conceived merely as the cultural inscription of meaning on a pregiven sex" because "gender is also the discursive/cultural means by which "sexed nature" or "a natural sex" is produced" (1990, 11). Thus, the Beauvoirian distinction between sex and gender is challenged because sex is shown to also be a social category, which is formed in the light of (rather than being a determinate of) gender categories (see Ásta 2018, 57-8).

This latter account of gender also does not hold that gender roles are necessarily so harmful and unwelcome. Thus, eradicating gender is not necessarily a goal. After all, as mentioned before, many gender-variant and transgender individuals enjoy expressing themselves with gendered roles

(Lev 2006, 46). Although Butler (1990) also maintains that gender is not a 'set identity' within the individual, it is still the chosen roles and pastimes which are performed by the individual, and so stem from them, which are then gendered in a gendered society. Other feminists have noted that women's genders can hold positive value for them, which would not disappear were gender to be eradicated and women were not to occupy a subordinate position in society (see Stone 2007; Mikkola 2016).

Now, the DSM-5 appears to employ the latter identity-based account of gender, as this is the only account with which criteria such as "an insistence that one *is* the other gender" (my emphasis) can make sense. This seems to rely on gender being self-generated and suggests that it is the expression of this inner identity with the relevant associated gender roles which fuels the preferences for and rejections of the gendered norms commonly associated with the sexes.

However, it is not clear how one would go about justifying that the DSM should indeed be using this identity-based account of gender in forming its diagnostic criteria for Gender Dysphoria (even if it is internally coherent to do so). The DSM may not be required to justify such things, but we may still more widely want to be able to justify why certain concepts and ideas about gender are used in this way to inform the categorization of mental disorder. But with reference to what? How should we choose between these accounts of gender in order to inform the classification of GD?

We are also still none the wiser with regards to what the link is between the dysfunction implicated in sex dysphoria and another dysfunction or the experience of gender role-dysphoria. Very little is understood about what dysfunction (if any) is present in cases of GD, when gender is understood as identity-based.

A traditional understanding of gender, describing gender as an external set of imposed social rules and expectations and therefore not as self-generated, would not be able to make sense of the idea of a dysfunction going on in what gender is projected onto you. This would have nothing to do with any natural mechanisms in the patient, functional or dysfunctional. The process of socialisation revolves around the treatment we receive from others, whether it be favourable or unfavourable depending on our sex. Understood as a social and cultural construct rather than a heritable and biologically evolved trait, it would be impossible to apply Wakefield's dysfunction analysis of natural mechanisms to this concept of gender (Bartlett et al 2000, 772).

So, depending on which understanding of gender we adopt, this significantly affects how we apply Wakefield's hybrid analysis of disorder and what phenomena we are then to look for. A dysfunction in forming a gender identity, or in coping with imposed gendered expectations? My aim here is merely to show the ramifications of this political question and the effects they have on attempts to use Wakefield's hybrid analysis to identify genuine mental disorder, and so I do not necessarily have to advocate for a particular one of these understandings of gender.

Lastly, with regards to sex dysphoria, the accounts differing on their ontological status of sex has ramifications for how this condition is understood. On a traditional account, we can indeed simply suffer from a misconceptualisation of what our physical bodies should look like, and which sex category we perceive ourselves as belonging to. On an identity-based account the picture isn't so clear, but one possibility is that if we conceptualise ourselves as belonging to some sex category and desire some surgical intervention, this can just be a reflection of the social engineering of sex categories which, when it doesn't follow normal expectations, indicates either a dysfunction somewhere or a state which is disvalued and pathologised.

4.3 Harm

So far, I have sketched out some key differences in two differing accounts of gender. On a more traditional view, sex determines gender in determining which gender roles are imposed on an individual, thus the sex and gender distinction is useful, and gender roles are harmful and should be eradicated. On the identity-based account, the performance of gendered activities categorizes someone as male or female, so sex is as socially engineered as gender and the sex and gender distinction breaks down. Finally, engaging in activities which happen to be gendered in society are what it means to have a certain gender, and these activities are not necessarily harmful. Which account of gender is adopted, has ramifications for how sex dysphoria is understood also.

I have not endorsed a particular account, but suggest ways in which these differences in the accounts of gender affect the identification of a dysfunction. It is not clear that these issues are just due to the requirement of context and value-judgements in identifying dysfunction, as is discussed by others (Lilienfeld and Marino 1995). Instead, I suggest that these issues are fundamentally sociological, with the matter of defining mental disorder intersecting head on with endeavors to understand gender and the mechanism of oppression.

One aspect which will be particularly pertinent to ascertaining whether harm (in Wakefield's sense of the term as stemming from disvalued dysfunction) is present in cases of GD is whether gender roles are inherently harmful or not. The two accounts of gender differ with regards to the nature of gender roles. According to the traditional view of gendered roles, these rules and expectations are *inherently* harmful. This is because they have been instilled into society at the expense of women's rights and freedoms and to the protection and furtherment of men's. According to the identity-based account, there is nothing inherently wrong or harmful about gender roles, but they only become problematic when an individual feels that those which are ordinarily applied to her are not appropriate for her. Finding gender roles harmful on a traditional account of gender would therefore be completely unsurprising. On an identity-based account, harm enters the picture when gendered behaviour is 'policed' and regulated by others, which would also be unsurprising.

However, Bartlett et al. discuss the difference in the nature of the harm being experienced with sex dysphoria and gender-role dysphoria, suggesting that "discomfort with one's biological sex and discomfort with the gender roles ascribed to this category are very different phenomena" (2000, 757). They provide evidence suggesting that much of the distress seen in children with gender-role discomfort can be traced to bullying, poor peer relations and their struggle against others' attempts to restrict their behaviours which are not seen as typical for their sex. Additionally, this distress is also often not at a clinical level. The distress of sex dysphoria, on the other hand, appears to be more directly caused by a dysfunction (Bartlett et al. 2000, 761-763).

Which account of gender we adopt affects *why* some identified dysfunction is experienced as a harm. This is something which a hybrid account of disorder doesn't take into account, but the reason why a dysfunction is harmful affects whether we want to say that the condition is disordered or just socially disvalued. This is more than just, on Wakefield's hybrid account, whether a dysfunction is present or not. Having some dysfunction may impede functioning and mean that you can't meet the cultural standards of a good quality of life, but it's important to ask why it has this effect. It may be for better or worse reasons. It might fail because the cultural standard for a good quality of life in place is good, and the condition in question just means that you can't meet it (for example, because it affects mobility, social connectedness, or causes chronic pain). Or, it might be that society is prejudiced and limits your quality of life when you have that condition. Why sex dysphoria is so harmful seems to be a case of the former; it's clearly very distressing and distracting to feel

that parts of your body are wrong and shouldn't be there. But it's not so clear with gender-role dysphoria and the rejection of certain gender roles why that is classed as a harm. Here, we see the hybrid account does nothing more than normative accounts do in evaluating *why* some condition is experienced as a harm, in order to try and avoid pathologising a socially disvalued natural state. Merely identifying a related dysfunction doesn't do this.

With an identity-based view of gender, it could be that the gender binary is insufficient when it comes to recognizing and accommodating the range of gender identities people have in society. With the traditional notion of gender, if we accept that the gender roles for women are inherently harmful then it would actually be expectable for women to reject those gender roles, seek more highly valued ones, and to be treated as the opposite sex etc. Although the criteria for GID in DSM-IV-TR included that GID cannot be "merely a desire for any perceived cultural advantages of being the other sex" (APA 2000, 581), it is not clear what we should use to base the difference between these two things on, and recognise each in any particular patient. Relatedly, the DSM-5 includes a brief discussion that 'gender non-conformity', which is when individuals behave, dress or have hobbies which do not match the gender norms of their assigned sex at birth, is different from GD and is not mental disorder (DSM-5, APA 2013, 458). However, again, it is not clear when cross-gender preferences *do* constitute symptoms of GD. The hybrid account fails to identify a useful dysfunction here to demarcate between gender non-conformity and GD.

It may be argued that cross-gender preferences constitute symptoms of GD when they are accompanied with serious clinical distress, but this could be greatly influenced by mere luck regarding whether the individual is surrounded by a progressive society and an open-minded family and peer group which accepts gender-variant behaviour. If one understands gender roles to be inherently harmful to women, then a significant amount of this distress could be attributed to the everyday enforcement of typical gender roles on women, and there may also be a matter of luck regarding how much freedom women may have in that environment. In fact, we do see an overrepresentation in women presenting to clinics and being diagnosed with GD, as well as an overrepresentation of those who have experienced trauma, are autistic, have pre-existing mental illness or are homosexual (Cretella 2017, 293).[2] As these conditions can also bring distress, it isn't

[2] Historically, though, boys were overrepresented in gender clinics. A discussion of this and why it might be so can be seen in Zucker et al. (1997). It is worth considering cases of men with gender-role dysphoria; on the traditional account of gender, despite gender roles being designed and instilled with

clear that we can attribute the harm and distress experienced by those diagnosed with GD solely to dysfunction, despite there seeming to be a dysfunction underlying sex dysphoria.

If we accept societal gender roles as inherently harmful, we may also be inclined to say that if those with gender-role dysphoria were taken away from this society with those harmful gender roles, then they would no longer be disordered. Yet, the definition of mental disorder in the DSM-5 states that "conflicts that are primarily between the individual and society are not mental disorders" and that they "must not be merely an expectable and culturally sanctioned response to a particular event" (APA 2013, 20). The removal of homosexuality from the DSM was largely motivated by the acceptance that gay individuals would live peacefully and without suffering in a world with no homophobia, because no harmful dysfunction was present. If gender roles vanished tomorrow, or certain pastimes were no longer disvalued for being feminine (and alternatively over-valued for being masculine), it may be that many individuals diagnosed with GD could live peacefully too. This is exactly the sort of pitfall which Wakefield claimed to avoid by bringing together both a normative and naturalistic component in an account of mental disorder, but simply linking one perceived harm to another perceived dysfunction in this instance has not been extensive enough to avoid beyond doubt pathologising a normal, expectable state.

In fact, the diagnostic criteria would not even be intelligible outside of a society, without any gender roles at all being present, because the criteria specifically refer to them. So, arguably, the very concept of GD could only emerge in a society with a widespread assumption that these gender norms are natural and inherent to the sexes, and can therefore act as markers of the 'true' gender of the individual rather than their sex or bodily constitution. *If* we were to accept a traditional account of gender, then this employment of gender roles in the criteria for a mental disorder reinforces them as natural and appropriate.

Of course, we might not accept the traditional account of gender. Importantly, as I previously noted, I do not necessarily need to endorse one of these accounts of gender here. The point is that on a traditional account of gender, we are pathologising a normal state, whereas with an identity-

the purpose of subjugating women, men can still suffer from this. Especially, those that are particularly uncomfortable with gender roles which relate to being bullish, independent, and emotionally detached. On the identity-based view of gender, men too experience isolation and social sanctions if they do not 'fall in line' with regards to expected gender expressions. Many thanks to an anonymous reviewer for raising these considerations.

based account this is not necessarily the case (there could be a disorder in the formation of one's gender identity). So, the matter of how gender should be understood has become relevant to whether we are accurately identifying a mental disorder in the case of GD. Wakefield's aim of identifying true disorder from merely disvalued states by bringing together a normative and naturalist element in an account of mental disorder is shown here not to be enough to do so satisfactorily. In investigating the specifics of dysfunction, harm, and the link between the two, we see the surprising extent to which a successful account of mental disorder will need to engage with sociological concepts and ideas, such as 'groups' in society, what a gender is, how gendered oppression works, to be able to define disorder.

Whether we endorse an identity-based account of gender or the traditional account of gender, we are still left with the question of what exactly is the nature of the link between on the one hand, sex dysphoria and a dysfunction based in body-conception, and on the other, gender-role dysphoria. This is the first shortcoming of the hybrid account; not investigating more closely how the harm and dysfunction relate to one another. I have shown how different understandings of gender affect whether dysfunctions can be coherently identified in sex dysphoria and/or gender role dysphoria. Perhaps, one of the reasons we were 'primed' to not recognise that it's not clear what the link is between sex dysphoria and gender-role dysphoria, might be just how pervasive and ubiquitous gendered expectations are in society. This means that we associate those gender roles so closely with the relevant sexes, that we don't wonder why one dysfunction should explain them both. The second shortcoming of the hybrid account I raise is not accounting for why some harmful dysfunction is experienced as harmful, even though a dysfunction may have already been identified. We need to identify harm which is caused by dysfunction, but also to be mindful of cultural influences on the construct of why that dysfunction makes life hard. In this case, according to a traditional account of gender, sexist notions of what pastimes men or women prefer, inform our decisions over the nature of the harm men or women may experience when they do not like them. On an identity-based account of gender, this could be an elusive dysfunction in the formation of a gender identity, or due to social disapproval when we engage in gender roles and pastimes which we are not expected to.

5. Appropriate Treatment

It is my view that GD should be removed from the DSM and not regarded as a disorder because there is no clear dysfunction (with either account of gender), but that sex dysphoria should remain. This is not so much due to endorsing some particular account of gender, but because it seems less likely that those with such intense discomfort with their sexed body, even from a young age, would cease to be disordered if they were placed in even an ideal social environment. Others, such as Giordano (2013) and Lev (2006) argue that GD in its entirety should be taken out of the DSM and not seen as a disorder at all, as the experiences associated with GD diagnoses are manifestations of individual differences in expression of gender and feelings about one's gender and/or sex, which should be seen as a natural part of human variation and do not cause harm and distress by themselves. Therefore, the classification in its entirety is mistaken in the same way that the classification of homosexuality was mistaken (and some of the detrimental repercussions of this may apply here). Giordano (2013, 55) argues there is no dysfunction present in the formation of gender identity in people who meet the criteria for GD, as there are no markers at all for 'ordered' and 'disordered' gender development. This would mean there is no harm due to a dysfunction.

She also argues that "gender and gender identity refer to the congruence between phenotype and the person's behaviour and feelings about oneself", or in other words, that gender identity is "the experience of belonging to a sex" (2013, 24). Therefore, Giordano maintains that one's gender and one's sex are fundamentally interlinked, such that someone who feels this incongruence, and that they should or do belong to the other sex, will also experience related desires and preferences to take on the roles and expectations usually associated with and considered usual for that sex within their social and cultural context. This would make it impossible for GD to be removed while sex dysphoria still remained in the DSM, and suggests a possible link between sex dysphoria and gender-role dysphoria. Perhaps that, once we start to feel that our gender role or our sex is inappropriate for us, that incongruence bleeds out into also affecting our comfort with the other.

Akin to Butler's (1990) ideas about cultural categories of gender forming the categories of sex, Giordano's link between gender and sex is that an individual's desires and pastimes interact with the culture's conceptions of male and female to form their gender identity and indicate which sex they feel a part of. This is how and why, in her view, our sense of our own gender can and does 'trump' whichever sex we are 'assigned'. Clearly, this

is in contrast with the traditional account of gender discussed earlier which defines gender roles as inherently harmful roles and expectations imposed onto female people. This conception of oppression is based on sex, whilst Giordano's appears to be based on gender identity.

Giordano maintains, similarly to myself, that the vast majority of distress suffered by those with less typical gender expressions is due to prejudice and marginalization, as we live in a society in which gender roles are rigorously enforced. However**,** I do not hold that this is the case for sex dysphoria also, and instead believe that sex dysphoria represents a harmful dysfunction that some individuals diagnosed with GD will have but others won't. As we have seen, some patients have symptoms which only relate to gender roles and other have symptoms which only relate to sex dysphoria, which raises questions about exactly when symptoms of one sort will and won't result in symptoms of the other sort, too.

Another issue with Giordano's view of GD and the link between gender roles and sex relates to effective treatment. The proposed treatments for GD include puberty-suppressing medications, cross-gender hormones or sexual reassignment surgery. These treatments are unusual in that they do not attempt to dispel and reduce the psychological symptoms of dysphoria, whether it be significant distress with one's gender role or one's physiological sex, but instead accommodate or affirm these symptoms (Meyer-Bahlburg 2009, 469). Giordano argues that this is perfectly acceptable on account of gender variant individuals not having a disorder and therefore not requiring treatment which dispels their symptoms without affirming them. Furthermore, this is in line with other treatments widely accepted to be appropriately administered by doctors despite the fact that they do not address a specific dysfunction, such as contraception or fertility treatment (Giordano 2013, 149-151). On (some) identity-based accounts of gender then, these treatments are aids in realising and manifesting to one's own satisfaction, one's own gender identity.

On other identity-based accounts of gender and the traditional account of gender, there may be concerns that such treatment fixes the individual in a way which 'gives in' to harmful and unideal societal norms and expectations, when perhaps it is the latter which should change.[3] It appears that we take a significant risk providing this nature of affirmative treatment when we do not have solid answers to the source of dysfunction and harm in some condition. In this case, we risk treatment being a way of

[3] Cretella raises the concern of appropriateness of affirmative treatment in other disorders which affect bodily conception such as anorexia, BDD or BID, because it's not clear that this type of treatment would be effective in reducing symptoms in the cases of those disorders (2017, 293)

reinforcing harmful gender roles in that we 'fix' the individual rather than society. Yet, Wakefield's hybrid analysis can be applied to the various understandings of gender with the various dysfunctions and harms which they posit, giving us no clearer a path for separating expectable states from disordered states. So, an accurate account of gender and the mechanism of gendered oppression is crucial also to ascertaining what type of treatment should be dispensed.

6. Conclusion

In this paper I discuss how different accounts of gender which vary on its ontological status, its distinction from sex, and whether it is inherently harmful, affect the identification of dysfunction and harm in some condition. Although I do not endorse here one account of gender or the other (there may well be complex accounts which incorporate elements from each account, such as Jenkins (2016)), I show that if we were to accept that sex is as culturally engineered as gender and so the distinction breaks down, this makes identifying the specific dysfunction in sex dysphoria difficult. If we accept a traditional account which posits sex as a biological category, a dysfunction in conceptualizing your sexed characteristics is more coherent.

With regards to gender-role dysphoria, the question of whether gender roles are understood as inherently harmful or not is pertinent. On a traditional view of gender, gender roles are inherently oppressive and marginalizing and so would naturally be experienced as harmful. On identity-based views of gender, someone could experience the harm of an elusive 'disordered' formation of gender identity, or more simply experience social ostracization for engaging in gendered activities which are not expected for them.

Wakefield's hybrid account doesn't consider how exactly the dysfunction and harm relate to each other, which would have highlighted the gap between sex dysphoria and gender-role dysphoria. It turns out that answering this question requires an entire account of sex and gender and how oppression on the basis of them occurs. It also doesn't consider, secondly, why the harm—even if it is related to a dysfunction—is experienced as harmful. This would give rise to questions about the nature of gender and sociology of oppression, and only then actually answer whether something is a disorder or not.

In this case, we identified harm which could have stemmed from inherently oppressive gender roles, or the marginalisation of gender variance (in presentation or self-identification), or from a dysfunction in the formation of a gender identity, along with possible dysfunctions in conceptualizing bodily constitution or in gender identity formation. Which to accept and how to relate them has been shown to be crucial in avoiding diagnosing healthy individuals with mental disorder. GD demonstrates the importance and relevance of the social theories we adopt and how they affect, to a surprising and up until now unacknowledged extent, whether or not we are pathologising individuals with normal or expectable mental states. My argument is quite reserved in that I do not suggest whether Wakefield's hybrid account of disorder can be updated or added to in a way which can address these concerns. Though, I suggest that similar concerns can be raised with regards to purely naturalist and normative accounts, and so will be a widely shared concern in defining mental disorder.

Acknowledgments

I would like to thank two anonymous referees for helpful and thoughtful comments on this paper. Also, to useful discussion which followed from presenting an earlier version of this paper to the meeting of Joint Session 2020 and to the Women in Philosophy group at the University of Birmingham. Finally, to Lisa Bortolotti for extensive discussion and feedback on these ideas.

REFERENCES

American Psychiatric Association and American Psychiatric Association, eds. 2013. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. 5th ed. Washington, D.C: American Psychiatric Association.

Ásta Kristjana Sveinsdóttir. 2018. *Categories We Live by: The Construction of Sex, Gender, Race, and Other Social Categories*. Studies in Feminist Philosophy. New York (N.Y.): Oxford University Press.

Bartlett, Nancy H., Paul L. Vasey, and William M. Bukowski. 2000. "Is Gender Identity Disorder in Children a Mental Disorder?" *Sex Roles* 43 (11/12): 753–85. https://doi.org/10.1023/A:1011004431889.

Beauvoir, Simone de. 1997. *The Second Sex*. Edited by Howard Madison Parshley. Vintage Classics. London: Vintage.

Bentler, Peter M., George A. Rekers, and Alexander C. Rosen. 1979. "Congruence of Childhood Sex Role Identity and Behaviour Disturbances." *Child: Care, Health and Development* 5 (4): 267–83. https://doi.org/10.1111/j.1365-2214.1979.tb00130.x.

Boorse, Christopher. 1975. "On the Distinction between Disease and Illness." *Philosophy & Public Affairs* 5 (1): 49–68.

Butler, Judith. 1990. *Gender Trouble*. New York/London: Routledge.

Cooper, Rachel. 2018. "Understanding the DSM-5: Stasis and Change." *History of Psychiatry* 29 (1): 49–65. https://doi.org/10.1177/0957154X17741783.

Cretella, Michelle. 2017. "Gender Dysphoria in Children." *Issues in Law & Medicine* 32 (2), 287–304.

Fairbairn, Catherine, Pyper Douglas, Gheera Manjit, and Philip Loft. 2020. "Gender Recognition and the Rights of Transgender People." *House of Commons Library*. https://commonslibrary.parliament.uk/research-briefings/cbp-8969/.

Gagné-Julien, Anne-Marie. 2021. "Wrongful Medicalization and Epistemic Injustice in Psychiatry: The Case of Premenstrual Dysphoric Disorder." *European Journal of Analytic Philosophy* 17 (2): (SI4)5-36. https://doi.org/10.31820/ejap.17.3.3.

Gatens, Moira. 2003. "Beauvoir and Biology: A Second Look." In *The Cambridge Companion to Simone de Beauvoir*, edited by Claudia Card, 1st ed., 266–85. Cambridge University Press. https://doi.org/10.1017/CCOL0521790964.014.

Giordano, Simona. 2013. *Children with Gender Identity Disorder: A Clinical, Ethical, and Legal Analysis*. Routledge. https://doi.org/10.4324/9780203097892.

Haslanger, Sally. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34 (1): 31–55.

James, Sandy E., Jody Herman, Mara Keisling, Lisa Mottet, and Ma'ayan Anafi. 2019. "2015 U.S. Transgender Survey (USTS): Version 1." ICPSR - Interuniversity Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR37229.V1.

Jenkins, Katharine. 2016. "Amelioration and Inclusion: Gender Identity and the Concept of *Woman*." *Ethics* 126 (2): 394–421. https://doi.org/10.1086/683535.

Jurjako, Marko. 2019. "Is Psychopathy a Harmful Dysfunction?" *Biology & Philosophy* 34 (5). https://doi.org/10.1007/s10539-018-9668-5

Lancellotta, Eugenia, and Lisa Bortolotti. 2020. "Delusions in the Two-Factor Theory: Pathological or Adaptive?" *European Journal of Analytic Philosophy* 16 (2): 37–57. https://doi.org/10.31820/ejap.16.2.2.

Langer, Susan J., and James I. Martin. 2004. "How Dresses Can Make You Mentally Ill: Examining Gender Identity Disorder in Children." *Child and Adolescent Social Work Journal* 21 (1): 5–23. https://doi.org/10.1023/B:CASW.0000012346.80025.f7.

Lawrence, Anne A. 2006. "Clinical and Theoretical Parallels between Desire for Limb Amputation and Gender Identity Disorder." *Archives of Sexual Behavior* 35 (3): 263–78. https://doi.org/10.1007/s10508-006-9026-6.

———. 2011. "Do Some Men Who Desire Sex Reassignment Have a Mental Disorder? Comment on Meyer-Bahlburg (2010)." *Archives of Sexual Behavior* 40 (4): 651–54. https://doi.org/10.1007/s10508-010-9720-2.

Lev, Arlene Istar. 2006. "Disordering Gender Identity: Gender Identity Disorder in the *DSM-IV-TR*." *Journal of Psychology & Human Sexuality* 17 (3–4): 35–69. https://doi.org/10.1300/J056v17n03_03.

Lilienfeld, Scott O., and Lori Marino. 1995. "Mental Disorder as a Roschian Concept: A Critique of Wakefield's 'Harmful Dysfunction' Analysis." *Journal of Abnormal Psychology* 104 (3): 411–20. https://doi.org/10.1037/0021-843X.104.3.411.

McKay, Ryan T., and Daniel C. Dennett. 2009. "The Evolution of Misbelief." *Behavioral and Brain Sciences* 32 (6): 493–510. https://doi.org/10.1017/S0140525X09990975.

Meyer-Bahlburg, Heino F. L. 2010. "From Mental Disorder to Iatrogenic Hypogonadism: Dilemmas in Conceptualizing Gender Identity Variants as Psychiatric Conditions." *Archives of Sexual Behavior* 39 (2): 461–76. https://doi.org/10.1007/s10508-009-9532-4.

Mikkola, Mari. 2016. *The Wrong of Injustice: Dehumanization and Its Role in Feminist Philosophy*. Studies in Feminist Philosophy. New York, NY: Oxford University Press.

Miyazono, Kengo. 2015. "Delusions as Harmful Malfunctioning Beliefs." *Consciousness and Cognition* 33: 561–73. https://doi.org/10.1016/j.concog.2014.10.008.

Moi, Toril. 2001. *What Is a Woman? And Other Essays*. Oxford: Oxford University Press.

Murphy, Dominic, and Robert L. Woolfolk. 2000. "The Harmful Dysfunction Analysis of Mental Disorder." *Philosophy, Psychiatry, & Psychology* 7 (4), 241–52.

Nicholson, Linda. 1994. "Interpreting Gender." *Signs* 20 (1): 79–105.

Nordenfelt, Lennart. 2007. "The Concepts of Health and Illness Revisited." *Medicine, Health Care and Philosophy* 10 (1): 5–10. https://doi.org/10.1007/s11019-006-9017-3.

Stegenga, Jacob. 2021. "Medicalization of Sexual Desire." *European Journal of Analytic Philosophy* 17 (2): (SI5)5-32.
https://doi.org/10.31820/ejap.17.3.4.

Stone, Alison. 2008. *An Introduction to Feminist Philosophy*. Cambridge: Polity Press.

Wakefield, Jerome C. 1992. "The Concept of Mental Disorder: On the Boundary between Biological Facts and Social Values." *American Psychologist* 47 (3): 373–88. https://doi.org/10.1037/0003-066X.47.3.373.

Wakefield, Jerome C., and Michael B. First. 2003. "Clarifying the Distinction between Disorder and Nondisorder: Confronting the Overdiagnosis (False Positives) Problems in DSM-V." In *Advancing DSM: Dilemmas in Psychiatric Diagnosis*, edited by Katherine A. Phillips, Michael B. First, and Harold Alan Pincus, 23–55. USA: American Psychiatric Association.: Advancing DSM: Dilemmas in psychiatric diagnosis.

———. 2012. "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): 130–39. https://doi.org/10.1016/j.comppsych.2011.03.001.

Wakefield, Jerome C., David Wasserman, and Jordan A. Conrad. 2020. "Neurodiversity, Autism, and Psychiatric Disability." In *The Oxford Handbook of Philosophy and Disability*, Edited by Adam Cureton and David T. Wasserman. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190622879.013.29.

Zucker, Kenneth J., Susan J. Bradley, and Mohammad Sanikhani. 1997. "Sex Differences in Referral Rates of Children with Gender Identity Disorder: Some Hypotheses." *Journal of Abnormal Child Psychology* 25 (3): 217–27. https://doi.org/10.1023/A:1025748032640.

European Journal of Analytic Philosophy EuJAP | Vol. 17 | No. 2 | 2021
https://doi.org/10.31820/ejap.17.3.6

THE QUANTITATIVE PROBLEM FOR THEORIES OF DYSFUNCTION AND DISEASE

Thomas Schramme[1]

[1] University of Liverpool

Original scientific article – Received: 22/03/2021 Accepted: 08/08/2021

ABSTRACT

Many biological functions allow for grades. For example, secretion of a specific hormone in an organism can be on a higher or lower level, compared to the same organism at another occasion or compared to other organisms. What levels of functioning constitute instances of dysfunction; where should we draw the line? This is the quantitative problem for theories of dysfunction and disease. I aim to defend a version of biological theories of dysfunction to tackle this problem. However, I will also allow evaluative considerations to enter into a theory of disease. My argument is based on a distinction between a biological and a clinical perspective. Disease, according to my reasoning, is restricted to instances that fall within the boundaries of biological dysfunctions. Responding to the quantitative problem does not require arbitrary decisions or social value-judgements. Hence, I argue for a non-arbitrary, fact-based method to address the quantitative problem. Still, not all biological dysfunctions are instances of disease. Adding a clinical perspective allows us to prevent the potential over-inclusiveness of the biological perspective, because it restricts the boundaries of disease even further.

Keywords: *theory of function; dysfunction; line-drawing problem; concept of disease; nosology*

Correspondence: t.schramme@liverpool.ac.uk

Introduction

Many biological functions allow for grades. For example, secretion of a specific hormone in an organism can be on a higher or lower level, compared to the same organism at another occasion or compared to other organisms. What levels of functioning constitute instances of dysfunction; where should we draw the line? This is the quantitative problem for theories of dysfunction and disease. It has increasingly been discussed in the philosophy of medicine in the past few years (Schwartz 2007; Hausman 2014; Griffiths and Matthewson 2016; Rogers and Walker 2017). Partly, the discussion is connected to the established debate between naturalism and normativism about the concept of disease. It seems that drawing boundaries between grades of normal and abnormal functioning involves value judgements, which undermine the naturalist ambition to devise a value-free theory of disease. In addition, the lack of a clear and widely accepted procedure for drawing the line seems to allow pathologisation of normal conditions as well as overdiagnosis (cf. Schramme 2019, 91ff.; Hofmann 2021). Every level of somewhat low organismic functioning seems to constitute a potential disease, if the line can only be drawn on the basis of human interests.

These practical concerns will form the backdrop of my contribution to the recent philosophical debate. I aim to defend a version of biological theories of dysfunction that exclude social value judgements. However, I will also allow evaluative considerations based on human interests to enter into theories of disease. My argument is based on a distinction between a biological and a clinical perspective (cf. Boorse 2014; Tresker 2020). The concept of disease, according to my reasoning, should be restricted to instances of biological dysfunctions. The use of 'should', in this context, implies that I do not believe in the possibility of pure conceptual analysis, resulting in a real definition of disease (cf. Lemoine 2013; Varga 2018). The best theory of disease will be determined by scientific considerations in combination with pragmatic interests, such as the avoidance of over-diagnosis.

Responding to the quantitative problem does not require arbitrary decisions or social value-judgements. Hence, I argue for a non-arbitrary, fact-based method to draw the boundary of dysfunction. Still, not all biological dysfunctions are instances of disease. Adding a clinical perspective allows us to prevent the potential over-inclusiveness of the biological perspective—in terms of potentially including too many diseases if we identify disease with biological dysfunction. To add a

clinical perspective helps to restrict the boundaries of disease to medically relevant dysfunctions.

In section 1, I introduce the quantitative problem within the context of a theory regarding the absolute concept of disease, that is, a conception that does not allow for grades of diseasedness. Section 2 briefly discusses the qualitative problem—which is concerned with identifying functions as opposed to non-functional mechanisms—in order to better understand the main concern of this paper. Only mechanisms that are identified as proper, performing functions are relevant for a theory of function and, derivatively, for a theory of disease. Hence, only functional traits are relevant for the quantitative problem. Section 3 then more thoroughly looks at the quantitative problem, specifically at Christopher Boorse's attempt to address it. I argue that this attempt struggles as it is, but can be repaired by adding clarity about the implications of seeing functions as effects within an organismic system. Thresholds for sufficient levels of functioning are determined in relation to next-level functions and the overall maintenance of the system. Accordingly, effectiveness of functioning is the relevant criterion for answering the quantitative problem, not the functional efficiency of a trait. In section 4, I draw a closer connection to medicine by introducing a perspective of clinical dysfunction, which is a narrower category than biological dysfunction. In section 5, I discuss the application of the general classification of clinical dysfunctions, which can be found in nosological systems, to individual patients through the process of medical diagnosis. Diagnosis therefore involves some discretion for clinicians when determining the boundary between normal functioning and pathology in individual cases. However, this practice is only possible within the boundaries set by the scientific notion of biological dysfunction. It therefore does not introduce wholly arbitrary elements. Section 6 concludes.

1. The Quantitative Problem in the Context of an Absolute Concept of Disease

In medicine, it is usually said that disease is the absence of health (in a specific respect, say, in respect to one's respiratory system). Health is deemed the opposite of disease. It is true that this conceptual binarity by itself does not establish clear-cut boundaries. Still, when we talk in this way, we interpret disease as an absolute concept. There are no grey areas; conditions either constitute a state of health or of disease. Things might be different when we consider whether a person is healthy, that is, when we consider health from a holistic perspective. From such a perspective, we

can easily consider conditions of disease that are consistent with a person being overall healthy.

From an absolute conceptual framework, we can also allow for positive health to be a gradable notion. We might consider healthier-than judgements, where one person is compared to another person (Schroeder 2013); yet these judgements do not result in grades of disease, because being less healthy is not the same as being unhealthy or more diseased. Disease is, so to speak, below the threshold of minimal health. It is true, of course, that different instantiations of diseases pose different levels of severity. Accordingly, we might want to say that a particular disease is more clearly a case of disease than another. Yet, if we have determined whether a condition is a disease, then it simply belongs to the class of disease, never mind how serious it is.

Such an absolute perspective is quite important in many practical contexts, most significantly when the presence of disease is used as a kind of entry ticket to the system of publicly funded medical resources. Here we need an absolute statement as to whether a condition is justifiably deemed a disease or not. If a condition is not a disease, it ought not to be treated by using publicly funded resources, at least not without additional argument. A condition that constitutes a disease, on the other hand, is a legitimate concern of a public health system without further reasons; although this might still not be enough to guarantee the public funding of treatment under the usual conditions of scarcity.

A serious problem for medical theory with respect to establishing an absolute concept of disease that has recently gained momentum is where exactly to draw the threshold between health and disease. What criteria need to be fulfilled in order to classify a condition as a disease? A common way to draw this boundary is to establish the criterion of dysfunction, or more exactly of impairment of functional ability (Boorse 1977). For the purposes of this paper, I will take such a Boorsean framework for granted, though I divert from Boorse in several respects. Accordingly, the general concept of disease is understood as impairment of functional ability. Functional ability is the readiness of a trait, for instance an organ, to 'pursue' its tasks. Accordingly, a trait currently not doing any work is still functional, perhaps due to environmental causes (Garson 2019, 126ff), if it has the relevant functional ability. Disease can therefore be understood as impairments of relevant dispositions within the Boorsean theoretical framework (Boorse 2014, 685). I will later identify disease with clinical dysfunction, which is based on, but not identical to, biological dysfunction.

Such a distinction is not thoroughly discussed in Boorse's theory, yet he explicitly allows for a clinical perspective on disease (Boorse 1997, 48).

There is an important difference between the theoretical problem of delineating disease as opposed to the same problem posed from a practical point of view. A doctor who deals with a suffering patient is not primarily interested in whether the organism in front of her is dysfunctional, but in her patient's wellbeing, broadly conceived. The doctor might therefore be tempted to identify a disease where there is no dysfunction or, conversely, not to diagnose a condition in terms of disease despite its being dysfunctional. It is important to disentangle different contexts of referring to dysfunction and disease, because they are based on different types of interests. I will distinguish between two such contexts: A biological and a clinical context.

From a theoretical point of view, aiming at an explanation of the concept of disease, the focus on the notion of dysfunction as a necessary criterion of disease allows us to establish an absolute concept of disease. Only where there is dysfunction, there can be disease. We are accordingly pushed back to the level of organismic functions and their impairments. However, individual organismic conditions and processes come in degrees. For instance, secretion of hormones allows for different values in different organisms at different times and under different environmental circumstances. Accordingly, when we focus on dysfunction as the basis of the concept of disease we seem to enter a grey area, after all, because the exact level of function that allows for a process to be called *dys*functional appears to be insurmountably vague. In other words, whether the concept of dysfunction allows for absolute thresholds and whether these can be established scientifically is not straightforward.

Some levels of performance can be deemed unambiguously dysfunctional, simply because they completely lack in functioning. Since a function is an effect of a trait, if a trait does not produce *any* such effect, it is dysfunctional. For instance, if a heart does not pump blood at all, it is dysfunctional. But surely there are many instances of organismic mechanisms producing effects that are however not sufficient to be deemed functioning. The problem discussed in this paper is how and where to draw this very line. I call it the *quantitative problem* of theories of dysfunction, because it is concerned with the level of producing an effect, not with the kind of effect a function is supposed to achieve.

Other authors have called the problem I will address “the line drawing problem”, most notably Peter Schwartz (2007), who was one of the first

authors to bring it up explicitly under this label in the philosophy of medicine, although there are important precursors to the recent debate (Engelhardt 1976; Goosens 1980). Boorse (1977, 1987, 1997) did attempt to tackle this problem in the past, from a naturalist point of view, but I believe there are problems with his account. Only quantifiable functions raise serious concerns where to draw the line; lack of any functional effect straightforwardly constitutes dysfunction. That is why I think the line drawing problem, which generally asks for the line between the functional and the dysfunctional, is in reality restricted to the quantitative problem. Accordingly, I prefer the latter label.

The term 'problem' is slightly ambiguous and it might be helpful to briefly explain in what sense I intend to tackle the quantitative problem. First, a problem can be something that is generally a matter of concern, for instance, especially in our context, a philosophical problem. In this way, a philosophical problem might never be solved; it might continue to be a matter of interest or concern, something that requires explaining. The mind-body problem might be a fitting illustration. It might never be solved and continues to interest us from a philosophical point of view. Second, a problem can also be something that bothers us in a certain way or that we want to get rid of. The mind-body problem might not be a problem in this second sense. Now, I believe the quantitative problem will continue to be a problem in the first sense of the term. It will continue to raise philosophical concerns. In this paper, I want to show a way to address the quantitative problem in a way that eases the problem in the second sense of the term, especially by showing a reasonable, non-arbitrary and workable way to conceptualise dysfunction and disease. I will not solve the problem in the first sense and will therefore avoid speaking of a solution to avoid any confusion.

In this way, I will defend an answer to the quantitative problem which claims to rely only on scientific aspects, hence avoids external evaluative elements, for instance in relation to individual harm, as other authors have introduced. The main idea is to identify the relevant level of gradable functioning with achieving a particular effect, relative to other functions of an organism. It is argued that the relevant threshold of quantitative functioning is determined by the biological necessities that are involved when a part of an organism, understood as an overarching system, is to perform its biological functions. Functions are effects, and any such effect is a means to maintain other functions, altogether maintaining the system as a whole. We can determine the required level of functioning in relation to the structured sub-systems of an organism. The quantitative problem therefore raises scientific questions regarding the biological organisation

of organisms. However, I will also argue that this only relates to the biological perspective. I suggest that in medicine we further need to account for a clinical perspective regarding the boundary between normal functioning and dysfunction. The clinical perspective introduces additional features, which are partly evaluative and pragmatic.

2. The Qualitative Problem

We can contrast the quantitative problem of dysfunction with the qualitative problem. The qualitative problem is concerned with identifying the kinds of traits of organisms that can be deemed functional, as opposed to being non-functional. Note that 'non-functional' means 'having no function'; it does not mean 'dysfunctional'. For instance, the function of the heart is to pump blood, not to produce noise, though the latter is also an effect of the organ's mechanisms. So, in other words, the qualitative problem aims at identifying the functions of traits. In the philosophy of biology, and also in the philosophy of medicine, this has been the major concern in the last decades. Several theories have been offered as to how to account for functions (a good range of papers can be found in Buller 1999 and Ariew et al. 2002; see also Garson 20016, for a helpful overview). I will not discuss these theories, because my main focus is on the quantitative problem.

To be sure, I do not want to deny that there is a close connection between the qualitative and the quantitative problem. After all, identifying functions (i.e. tackling the qualitative problem) usually comes with specific quantitative levels of functioning (see Schwartz 2007, 366). So, for instance, the heart does not simply have the function to pump blood but to pump about 5 litres per minute in a resting adult person. I still want to insist on the difference between the two problems for analytic purposes, because later I will argue that examples of quantifiable non-functional traits do not apply to the quantitative problems. In other words, only functional traits raise the relevant problem. In general, it seems to me that the second aspect of the qualitative problem—specifying functions over and above identifying traits with functions—can be translated into the quantitative problem, because it causes the need to determine thresholds for normal functioning.

Many theories of function can account for dysfunction or malfunction. This is usually, though not universally, done by using the type-token distinction (Godfrey-Smith 1993, 200). More specifically, types of traits are explained to have specific functions. In the medical context, 'trait' may

stand for organismic sub-systems, such as the respiratory system, organs, cells, or even genes. All of these things may have a function, and it is of course an important problem for biological and medical research to explore these functions. Different theories of functions differ in their explanations as to why a particular effect is the function of a trait. It might be due to its evolutionary history (Wright 1973; Millikan 1989; Neander 1991), its contribution to a larger system's capacities (Cummins 1975), or to the good of the organism (Melander 1997; McLaughlin 2001; Wouters 2003). Once a type of a functional trait is established, tokens can be assessed according to the norm set by the functional type. Accordingly, the qualitative problem regarding dysfunction is concerned with identifying those features, or qualities, of organisms that can be dysfunctional at all. If a trait does not have a function, it cannot be *dys*functional.

The qualitative problem also addresses the problem of how a trait can be dysfunctional. Once the function of a trait is established, we know in what way a token can be dysfunctional, namely in terms of the effect that is its function, not by lacking in terms of other effects. For instance, a heart can be dysfunctional in terms of blood-pumping, not in terms of noise-production.

This is all I will say about the qualitative problem. It should nevertheless be pointed out that many issues in relation to the qualitative problem have not been sufficiently tackled in the philosophy of biology and the philosophy of medicine, for instance the related problem whether proper functions come in degrees (Matthewson 2020). Most notably, the specific normativity of function statements, which is supposed to account for the possibility of dysfunction, or malfunction, is also still a contested issue (Neander 1995; Davies 2001; Garson 2019).

3. Taking the Sting Out of the Quantitative Problem

As I have said already, the quantitative problem regarding dysfunction is due to the fact that many functions allow for degrees. At least in some contexts—especially where we need to determine unequivocally whether a condition is pathological—it seems to require an element of human decision. This itself does not need to be dubious, but rather normal procedure in relation to vague terms. In the philosophical debate, where to draw the line is normally regarded a problem for two distinguishable reasons: First, because it might involve an element of value judgement. This would threaten specifically the ambitions of a naturalist account of disease (Miller Brown 1985, 5f.; but cf. Veit 2021; Amoretti and Lalumera

2021). Second, line drawing could be problematic because it might not allow for any answer that is reasonable, non-arbitrary and "workable" (Schwartz 2017, 495). I will mainly focus on the second interpretation of the line-drawing problem and suggest a scientific response. The first interpretation of the problem requires further considerations regarding what types of value judgements are involved when drawing the line between dysfunction and normal function. Although I cannot go into detail here, I believe that any evaluations that might be involved will not be based on individual or social value judgements (Schramme 2010), but refer to the natural normativity of biological functions; hence be grounded on a scientific explanation of abnormality (Matthewson and Griffiths 2017, 452).

Interestingly, some organismic functions are not affected by the quantitative problem, at least they might relatively easily allow for non-arbitrary and workable results. There are some effects that only allow for absolute levels of performance. For instance, a two-way switch is either fulfilling its function or dysfunctional, depending on whether it can be turned or not. There are similar kinds of mechanisms in human organisms where the threshold of dysfunction is straightforward, even if the function does allow for grades. A function of the ovaries is, for instance, to produce eggs. At least, this is a function of the ovaries during a particular period of the life of a female organism. If a token ovary does not produce eggs, it is dysfunctional in that respect. So, in cases such as the one just mentioned, the way to tackle the quantitative problem is straightforward.

My example seems to raise concerns, however, that lead us into less straightforward terrain: After all, isn't the function of the ovaries to produce fertile eggs, at the right time, as well as, probably, to only produce one egg within one cycle? Being fertile seems to clearly allow for gradual aspects, for instance regarding how likely an ovum is to develop into a zygote, once fertilised. How fertilisable an ovum is clearly depends on numerous other functions and environmental conditions. My example was mainly meant to establish the possibility of isolated functions, where thresholds are relatively easy to determine, not to exclude more complicated gradual functions. If ovaries do not produce eggs—never mind any gradable characteristics of these—then they are dysfunctional. If they produce more than one ovum during a cycle, they might also be dysfunctional. Whether this is the case or not does not matter for my purpose, as I will agree that additional, non-biological considerations are required for determining dysfunctions in a clinical context.

The quantitative problem is more difficult to tackle when traits allow for different levels of performance without clear thresholds. The most common way to provide a threshold, at least in abstract terms, is to say that a trait is dysfunctional if it does not perform efficiently. In the philosophy of medicine, Boorse, who defines disease as impairment of functional ability, has described the threshold of dysfunction in the following way:

> Normal functioning in a member of the reference class is the performance by each internal part of all its statistically typical functions with at least statistically typical efficiency, i.e. at efficiency levels within or above some chosen central region of their population distribution. (Boorse 1977, 558f.)

The important point here is to be found in the final part of the sentence. Boorse makes clear that he wants to account for the threshold by statistical means. There are, however, serious problems with such a framework (Schwartz 2007; see also Davies 2001, 186).

As mentioned before, several authors have objected that a statistical answer to the line-drawing problem is arbitrary. If this is true, it might, firstly, show that dysfunction cannot after all be explained in purely value-neutral, scientific terms. The quantitative problem might require reference to particular human interests, which Boorse would like to exclude from his theory of disease. Secondly, his theory leads to problems with low levels of functioning that are prevalent in a population. A statistical analysis does not work if inefficient levels are statistically normal. I will briefly deal with both of these objections, before presenting an alternative answer to the quantitative problem. I consider my considerations to apply within a generally Boorsean theoretical outlook. There might be alternative theories of function, for instance Cummins-style systemic theories of function (Cummins 1975) that fare better with the quantitative problem. The main purpose of my paper is, however, to present an alternative reply based on Boorse's framework.

Boorse himself maintains, in the quoted sentence, that the exact boundary between efficient function and dysfunction is "chosen", i.e. determined by human choice. However, he insists that the chosen region within the population distribution, which is deemed to be below the efficient level of functioning, is not chosen for reasons of human welfare interests, or the like, but for reasons of statistical theory. He says that 'deficiency', according to his account, is an "arithmetic, not an evaluative, concept" (Boorse 1997, 21). This might be so, but it nevertheless introduces an element of human decision about where to draw the line of pathological

levels of functioning. Indeed, Boorse himself says that "the lower limit of normal functional ability—the line between normal and pathological—is arbitrary" (Boorse 1987, 371).

Here the second problem looms, as there are some low levels of performance, which are so common that they will never stick out statistically. A common way to deal with this problem has been to put the relevant functions in relation to normal environments (Hausman 2014). Boorse addressed the related problem of statistically common or universal diseases, such as caries, already in his early papers. He also laid out a theory that refers to an environmental clause, so that environmentally caused or sustained dysfunctions are not deemed diseases (Boorse 1977, 566ff.). All these fixes seem to lead to the conclusion that normal levels of functioning cannot wholly be determined intrinsically, that is, only by reference to the organism and its internal mechanisms itself. This might not be devastating, but nevertheless, to include normal environments in the definition of normal functioning simply shifts the problem as to where the threshold of abnormality lies from one aspect to a perhaps even more contested one (see also Kingma 2010).

It is hard to deny that the exact level of performance needed in order to fall within the area of normal functioning is difficult to draw and indeed vague. The reference to statistics makes this even more evident. After all, there are no logical or conceptual reasons to see any normative significance in, say, the fact of two standard deviations in any measured value. Boorse says: "whenever one knows the goal of a process, one knows what is more or less function, and 'deficiency', in the context quoted, simply means much less than average" (Boorse 1997, 21). For Boorse, "functional efficiency [is] measurable" (Boorse 2014, 690; see also Kraemer 2013) and the boundary to dysfunction is due to a significant distance from the statistically determined mean level of functioning. Yet to concede that there is necessarily an element of human choice involved actually underlines the point which critics have brought forward. Critics say that this feature, the element of human choice, challenges Boorse's claim of providing a value-free theory of disease (Schwartz 2007; Kingma 2010).

However, I believe Boorse's reliance on statistics is wrongly conceived. Statistics is only an instrument to gain knowledge about organisms, not itself the source of drawing the line between function and dysfunction. It is important to see that the ontological perspective on the boundary between function and dysfunction is different from an epistemological perspective. The ontological perspective has to do with the level of performance of a type of trait; the epistemological perspective is required

to gain knowledge about the required level of functioning (cf. Hauswald and Keuck 2017).

The fact that we are here referring to types should lead us to acknowledge that setting the ontological boundaries requires a certain amount of abstraction and idealization. Surely the level of normal performance of a mechanism is not straightforward. Still, in cases of organisms that are structured through different levels of sub-systems the thresholds are determined by the relevant effect that is minimally required to maintain the relevant subsystem altogether. This is mainly, although probably not exclusively (because some environmental factors might need to be acknowledged), due to characteristics of the type of organism itself.

The quantitative level of normal performance for an organismic mechanism or process is determined by the requirement of achieving the effect that is its function. In other words, we need to see the performance of a trait as a means to an end (McLaughlin 2009, 96ff.). The end of a functional mechanism is a particular effect. Any level that achieves the effect is normal; any level of performance—high or low—that fails to maintain or to lead to the required effect is dysfunctional. Hence, the threshold of functional efficiency is determined by specific effects of biological processes.

So far, my argument seems to be circular: The line identifying the level of efficient functioning is drawn by the function of a trait. A trait is functional if it fulfils its function; dysfunctional if it does not fulfil its function. However, the required effects are themselves to be seen in relation to the hierarchical organisation of organisms. An effect is needed, usually together with other effects, to maintain functioning on a more complex level. Hence, effects (i.e. functions) of a trait are means to other ends. For instance, a function such as hearing requires many sub-functions being achieved. A heart needs to fulfil its functions to maintain other systems in the organism. It is not arbitrary or unworkable to determine the amount of blood pumping to achieve these other effects.

It needs to be stressed again that my suggested response to the quantitative problem is still not a solution to the problem as such, in the sense of getting rid of it once and for all. I nevertheless hope to show that this is not damaging, because at least my response opens a way of identifying reasonable, non-arbitrary and workable thresholds. It is true that the effects (functions) of connected organismic systems, which are supposed to determine the level of functions in maintaining traits, are themselves usually gradable. Hence the quantitative problem does not dissolve. For

instance, the threshold of cardiac output—say, 5 litres per minute—is effective in relation to a gradable performance of the organism. Now, we need to know what level of organismic performance we are using as baseline. The relevant effect might be required in a state of rest, whilst running, or in any other possible conditions of an organism. Yet, once we have settled on the respective effects of an overall system, due to our research interest, what level of functioning is required for maintaining the systemic functions is a matter of fact.

Biology can account for several functional systems within a type of organism and for their interdependence (Saborido et al. 2016). At least in biological theory, the ontological perspective on drawing the quantitative boundary between normal levels of performance and dysfunction—even where there are grades of performance—can be addressed by purely factual considerations. After all, the exact level of required or normal performance is determined by the factual question as to whether a particular effect can (still) be achieved. It should be added that this also allows for compensatory mechanisms to take over a function or making up for quantitative loss (Saborido et al. 2016, 113).

Boorse himself had stated a similar idea in an early essay:

> In fact, the structure of organisms shows a means-end hierarchy with goal-directedness at every level. Individual cells are goal-directed to manufacturing certain compounds; by doing so they contribute to higher-level goals like muscle contraction; these goals contribute to overt behavior like web-spinning, nest-building, or prey-catching; overt behavior contributes to such goals as individual and species survival and reproduction. What I suggest is that the function of any part or process, for the biologist, is its ultimate contribution to certain goals at the apex of the hierarchy. (Boorse 1977, 556)

I do not believe that we need to endorse Boorse's idea of a hierarchy of functions including an apex. In other words, we do not need to assume that biological systems have overall purposes, such as survival or reproduction in case of Boorse's theory. This assumption has raised numerous concerns (Cooper 2002). It is sufficient to agree with the interpretation of organisms as a conglomerate of subsystems involving functions on different levels—where *levels* is meant as a spatial term.

The way I have interpreted the quantitative problem makes clear that the relevant concept is not, as Boorse has claimed, functional efficiency, but

rather functional *effectivity*. In contrast to efficiency, the notion of effectivity comes with an internal absolute threshold, namely whether a specific effect is reached or not. In relation to the threshold a level of performance is either effective or not. In this reading there is no grey area. This shows, to my mind, that the suggested answer to the quantitative problem is in congruence with the straightforward cases of complete failure of any level of function. After all, not reaching the effect, which is the function of a trait, is simply failure of relevant performance.

It is true, of course, that we can introduce gradual interpretations of functioning, for instance regarding hearing. A person might be able to hear better or worse. I am suggesting, though, that once we will have determined an ideal type of the functional system for human hearing, we can decide whether the person's hearing is dysfunctional without considering its comparative level of overall performance. Normal hearing will be understood as a set of functions performing effectively on different interlinking levels. These functions will, at least for eventual clinical purposes, need to be modelled relative to age in order to produce reasonable thresholds that take senescence into account. To be sure, some of these functions will be set by quantitative measures, but the quantitative threshold levels will be determined by the respective required effects to maintain a system of functions. In other words, the thresholds of quantitative functioning will be set by the necessities of maintaining an organismic system. These are determined by an idealised model of a type of organism, relative to certain additional features, such as age or sex. Such an idealised model is the product of humans, of course. But it is not based on unreasonable, arbitrary or unworkable assumptions.

The ontological perspective, I have said earlier, is different from the epistemological perspective. Indeed, it is obvious that it is not easy at all in practice to establish the exact boundaries of normal function and dysfunction, though we have a theoretical instrument in modelling an ideal type of a functioning organism. I believe this is where statistical considerations can be of some importance (see also Hausman 2014; Garson and Piccinini 2014, 10ff.). After all, we cannot simply read quantitative values of normal performance off nature, but need to determine them by studying real specimens of the relevant organisms. Hence, we might use statistics as a means to gain knowledge about these levels of normal functioning. Yet we should now be able to see that statistics only provides clues for supporting certain theoretical assumptions about the ontological threshold between normal functioning and dysfunction. Statistics cannot itself *establish* the ontological boundaries, because the latter are determined by biological facts. We have seen already that statistics might

also lead to epistemic problems in cases of endemic malfunction or universal diseases. However, statistics is not our only means of gaining knowledge about functions. In biology, reverse engineering, for instance, is a common mode of developing models of the functioning of organisms (Smith 1995, 3; Green 2018).

4. Biological and Clinical Dysfunction

So far, I have discussed the quantitative problem as a classificatory issue within biology. And it is such a problem, of course. We want to know where to draw the boundaries between normal function and dysfunction, and this task need to be performed in relation to the organisms we study. Hence, we focus on the biological features of a specific type of organism to establish a prototype. But the quantitative problem in medicine is not only a biological issue; it is a clinical issue as well. We need additional considerations in this perspective.

I started by pointing out the normative significance of the threshold between normal function and dysfunction for calling a condition a disease. Most importantly, the boundary has an impact on people's access to publicly funded healthcare resources. Medical classification relies on a theory of the abnormal functioning of a specific type of organism, so the clinical perspective builds on biological considerations. A medical nosology for human beings, for instance the *International Classification of Diseases* (ICD), builds on a kind of normative prototype of a healthy human being—or rather it gathers several prototypes of abnormal functioning, specified according to different systems of organismic organisation. The ICD is organised along diseases of the blood, of the immune system, of metabolism, the nervous system, the visual system, and so on. But classification for clinical purposes does not stop at biological considerations. Clinical prototypes already contain pragmatic elements, which have to do with non-biological aspects, such as whether a condition can be identified or treated by medical means and has any impact on human wellbeing (see Cooper 2020, 154f.).

To be sure, we can imagine a medical nosology that rests exclusively on a biological foundation. Medical terminology indeed contains the term "subclinical", which might at least partially account for a purely biological perspective. I assume that the notion of the subclinical is actually intended to record early stages of processes that might (very likely) result in disease, though are not themselves instances of disease. Similarly, a purely biological classification would serve the purpose of recording any known

organismic dysfunction. This might be a relevant purpose for medicine as a scientific endeavour. But surely the purpose of such a system would be purely biological, not clinical. Indeed, to merge the biological and the clinical perspective easily lends itself to the problem of overdiagnosis. There are many biological dysfunctions that will not have serious effects on overall organismic functioning in any token organism, especially at a more microscopic level of the functional system. To call all biological dysfunctions diseases can have serious practical consequences because of the normative effects that usually come with the use of a disease label.

In this context, it has been said that Boorse's theory of disease is overly inclusive, because it rests on dysfunction, and dysfunction can surely be present on a cellular level. Hence even "one dead cell" would be pathological, according to Boorse's theory, which seems counterintuitive (Nordenfelt 1995, 28; Wakefield 2014, 656; Doust et al. 2017). Boorse himself has responded to this objection by accepting the implication and maintaining that every person has at any time some pathological condition, if only very minor, of course (Boorse 1997, 50f., 85; see also Boorse 2014, 706; 2015). But I believe we can respond to the charge of over-inclusiveness by pointing out that the classification of dysfunctions for clinical purposes, i.e. the classification of diseases, is different from a purely biological classification of dysfunction. It might be true that everyone has at any given time a biological dysfunction present in their organism. But the concept of dysfunction for clinical purposes adds further criteria to eventually result in the concept of disease.

I doubt that these additional criteria are convincingly understood by simply adding a harm condition, as some authors, most notably Jerry Wakefield, would like to convince us (Wakefield 1992; 2014). Clinical classification serves several aims, which I cannot thoroughly discuss in this paper (for an interesting analysis from a historical perspective, see Jutel 2011). There seem to be numerous examples of clinical diseases (listed in the ICD-11) that are not themselves harmful, say, for instance, benign skin tags (code EK 71.0) or protruding ears (code LA 21.1.). These conditions usually are considered for medical treatment, that is, they in fact qualify as entry ticket to use publicly funded resources. For the present purposes, however, my objection to an added harm criterion is not particularly relevant. It is more important to point out that although biological considerations build the basis of medical nosology, biological dysfunction is not sufficient for disease from a clinical perspective. We need a clinical understanding of dysfunction as well. One dead cell will not be seen as pathological from a clinical perspective.

I should stress that these additional criteria for clinical purposes bear on the quantitative problem of drawing the boundary between normal function and dysfunction. Although biological dysfunction is based on factual aspects regarding traits not achieving their supposed effects, the concept of clinical dysfunction is not merely based on factual aspects. Still, I want to argue that the additional considerations for clinical purposes do not undermine the foundational factual elements of biological dysfunction.

An example that has been discussed to show that Boorse's account has problems with drawing the line between disease and health is hypertension (Rogers and Walker 2017, 410). The exact line of a pathologically high blood pressure seems arbitrary, in other words not factual at all. How would this example pan out in the account I have introduced? It would need to be checked what quantitative value of blood pressure, if any, typically goes along with a lack of achieving the effects of related functions of the vascular system. As I have said earlier, we would need to abstract from individual cases and devise a normative prototype of normal blood pressure. Now, the specific example might appear not be pertinent, anyway, because it seems that blood pressure itself is not a functional feature of organisms, but merely a symptom of possible dysfunctions, especially of future dysfunctions (see Hofmann 2021, 131). Still, quantitative levels of blood pressure are indications of levels of functioning. Very high values of blood pressure are causally associated with pathological conditions, especially heart and kidney diseases. To be sure, this is a statistical correlation, indicating a specific risk of disease, not disease itself. In some cases, abnormal blood pressure might be a sign of a dysfunction, but again hypertension itself would not constitute dysfunction. Altogether, blood pressure is not a straightforward example of a functional trait. It is not clear whether it poses specific problems for a scientific theory of disease, because the quantitative threshold would be set by the requirements of maintaining the relevant organismic system.

Additionally, within the abnormal range, we might want to further enquire, from a clinical perspective, whether all subnormal levels are posing risks for human wellbeing or affect any other additional criteria. Still, these additional considerations would only be pursued below the threshold set by biological considerations. In other words, only biological dysfunctions would qualify as clinical dysfunctions. Hence there is no special danger of including too many conditions as diseases, in other words, no concern of pathologisation or overdiagnosis.

To be sure, I have only discussed one example that was used in the scholarly literature to establish the arbitrary nature of attempts to tackle the

quantitative problem. There might be other, more pertinent cases, which could undermine my claim that biological dysfunction sets the boundaries for determining clinical dysfunction. But as long as these can be accommodated, my claim regarding the scientific boundaries of clinical dysfunction still stands.

In summary, I have argued that the quantitative threshold lies where the specific effect, which is a trait's function, cannot be achieved or maintained. This relates to the biological notion of dysfunction. In a clinical context, there will be additional considerations. Still, these need to be based on the biological account. There can only be pathology where there is biological dysfunction (Matthewson and Griffiths 2017, 449; cf. Hucklenbroich 2017). Not every biological dysfunction is necessarily a case of clinical dysfunction, as we have seen when briefly discussing the "one dead cell" problem.

It is easily imaginable that we will have different quantitative measures for clinical purposes, which are more lenient, as it were. For instance, any value of myopia might be dysfunctional from a biological point of view, at least if we disregard aspects of normal deterioration of eyesight due to senescence for the time being. After all, the very notion of myopia seems to be based on an assessment of a trait as dysfunctional. The effect of sharp representation of an image on the retina is not achieved if an organism has myopia. However, clinically speaking it is likely that we will accept minor levels of myopia within the normative prototype, perhaps because perfect eyesight is so rare or because it normally does not bother people. Accordingly, there are external values and human interests involved when drawing the boundary to those biological dysfunctions that are clinically pathological.

Similarly, in psychiatry it is common to include in the classification of several disorders a clause that a specific condition must be present for more than six months. From a biological point of view, if a mental dysfunction is present, it will be present at any point in time, not just after some period of time. To be sure, we might use the time factor for epistemic reasons, in order to gain sufficient knowledge about the actual mechanism and whether it is still functioning in an individual person. But be that as it may, the biological and the clinical perspective can fall apart, simply because the relevant thresholds can differ (see Cooper 2013).

The fact that clinical considerations are partly driven by human interests, mainly by considerations of the impact of a biological dysfunction on wellbeing or the possibility of treatment, should not conceal the other fact

that within this perspective the concept of disease still has a firm scientific basis in biological dysfunction. Only biological dysfunctions can be deemed diseases, though not all will. This is different from accounts of disease that start from a social-evaluative foundation. The account developed here helps avoiding pathologisation of normal conditions and can be instrumental in preventing overdiagnosis. Admittedly, the latter achievement depends on the characteristics of non-scientific elements used in actual medical practice. It is true that in many countries medicine tends to cater for ever more minor biological dysfunctions and even for other conditions that are not biological dysfunctions at all. But this problem is a political one and scientific theories of disease cannot be blamed for it.

5. The Role of Diagnosis

So far, I have discussed the quantitative problem in relation to what I have called normative prototypes, hence on a generic level. It is a problem for medical nosology. But assessments regarding dysfunction in clinical medicine are also made on an individual level. Doctors make statements about individual tokens of organisms, also known as patients. These medical judgements are called diagnoses. The process of medical diagnosis leads to further complications for the quantitative problem of the boundary between normal functioning and dysfunction, because it opens some space for individual deviation from a normative prototype. The specific situation of a patient, who is of course not merely regarded as an organism when presented to a doctor, partly drives the assessment of functional capacity. A condition that is clearly clinically dysfunctional and hence pathological according to the relevant classification might not be diagnosed thus by a doctor. It might happen that an individual will not be subsumed under a prototype despite fulfilling the criteria of inclusion.

In terms of the quantitative problem the flexibility for diagnosis might work both ways; that is, there might be a diagnosis of a pathological condition where the individual patient is within the area of clinically normal functioning. For instance, for professional sharpshooters even the slightest level of myopia might have devastating effects on their career. Accordingly, a doctor might diagnose a relevant pathology. Note that this is different from diagnosing an alleged disease outside the range of biological dysfunction. A sharpshooter might prefer to have a vision comparable to an eagle; but biologically normal levels of human functioning can never be diseases within the suggested theory.

There might also be reasons for a doctor to avoid diagnosing a disease although the person presents with a clinically abnormal value. For instance, a teenager with extremely tall parents might have a growth hormone dysfunction, leading to stunted growth, but also resulting in a predicted height that is statistically normal. In such a case, it does not seem required to diagnose a disease. Such scope for deviance from clinical classification is actually desirable, because clinical classification by its very nature cannot account for individual cases. Yet, in medical practice it is important to do justice to individual cases.

It should also be stressed that any judgement regarding disease in individual cases is due to a diagnostic process. Perhaps in contrast to common expectation, the presence of disease is never fully established by pathological findings alone—which might for instance be achieved by investigating samples of tissue. Diagnoses are made by specialist doctors in relation to a patient. Their verdict is of course informed by pathologists' reports, but an individual judgement regarding disease within the clinical context is not merely due to a finding of clinical dysfunction. Admittedly, it seems that this practice is changing in reality and doctors tend to look more at laboratory results than at the patient to draw a diagnosis. But this development actually undermines the significant difference between the biological and the clinical perspective and should therefore be criticised.

If medical nosology could always sufficiently determine whether an individual case falls under a type of disease, the exercise of diagnosis would be merely deductive. Potentially a computer could then do the diagnosis, because the only question would be whether a person, or case for that matter, presents certain conditions, which form the criteria of a specific concept of disease, defined in a classificatory system, such as the ICD. But diagnosis is not simply a deductive exercise, and it should not be. Surely this aspect of the quantitative problem involves evaluative considerations that transcend the mainly scientific or factual aspects I have discussed above (cf. Whitbeck 1981).

6. Conclusion

In this paper, I have defended a particular way of accounting for the boundary between normal biological functioning and dysfunction. I claim that this boundary is due to matters of fact, yet not constituted by statistical realities. The quantitative problem can be dealt with in a non-arbitrary way. Functions are specific effects, which are either achieved or not. This is a factual question about the quantitative necessities to perform biological

functions within a complex structured organism. In virtue of exploring the systems of organismic functioning, biology develops normative prototypes, which can be used for medical purposes. However, when switching to a clinical perspective, additional considerations are introduced. Hence biological dysfunction is not the same as disease. Matters become even more complicated with individual diagnoses, which establish the reality of instances of disease in the actual practice of medicine. Medical diagnosis requires a judgement that puts clinical types and individual persons in relation.

Acknowledgments

I would like to thank Dan Hausman, Ulrich Krohs, Steve McLeod, Wendy Rogers, Peter Schwartz and two anonymous reviewers for their helpful comments.

REFERENCES

Amoretti, M. Cristina, and Elisabetta Lalumera. 2021. "Wherein Is the Concept of Disease Normative? From Weak Normativity to Value-Conscious Naturalism." *Medicine, Health Care and Philosophy*, August. https://doi.org/10.1007/s11019-021-10048-x.

Ariew, André, Robert Cummins, and Mark Perlman. 2002. *Functions: New Essays in the Philosophy of Psychology and Biology*. Oxford University Press.

Boorse, Christopher. 1977. "Health as a Theoretical Concept." *Philosophy of Science* 44 (4): 542–73. https://doi.org/10.1086/288768.

———. 1987. "Concepts of Health." In *Health Care Ethics: An Introduction*, edited by Donald Van De Veer and Tom Regan, 359–93. Temple University Press.

———. 1997. "A Rebuttal on Health." In *What Is Disease?,* edited by James M. Humber and Robert F. Almeder, 3–134. Biomedical Ethics Reviews. Totowa, NJ: Humana Press. https://doi.org/10.1007/978-1-59259-451-1_1.

———.2014. "A Second Rebuttal on Health." *Journal of Medicine and Philosophy* 39 (6): 683–724.

———. 2015. "Reply to Wakefield on Harm and Health." *MS.*

Brown, Miller W. 1985. "On Defining 'Disease'." *Journal of Medicine and Philosophy* 10 (4): 311–28. https://doi.org/10.1093/jmp/10.4.311.

Cooper, I—Rachel. 2020. "The Concept of Disorder Revisited: Robustly Value-Laden Despite Change." *Aristotelian Society Supplementary Volume* 94 (1): 141–61. https://doi.org/10.1093/arisup/akaa010.

Cooper, Rachel. 2002. "Disease." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 33 (2): 263–82. https://doi.org/10.1016/S0039-3681(02)00018-3.

———. 2013. "Avoiding False Positives: Zones of Rarity, the Threshold Problem, and the DSM Clinical Significance Criterion." *The Canadian Journal of Psychiatry* 58 (11): 606–11. https://doi.org/10.1177/070674371305801105.

Cummins, Robert. 1975. "Functional Analysis." *Journal of Philosophy* 72, 741–65.

Davies, Paul Sheldon. 2001. *Norms of Nature: Naturalism and the Nature of Functions*. A Bradford Book. Cambridge, Massachusetts: MIT Press.

Doust, Jenny, Mary Jean Walker, and Wendy A. Rogers. 2017. "Current Dilemmas in Defining the Boundaries of Disease." *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine* 42 (4): 350–66. https://doi.org/10.1093/jmp/jhx009.

Engelhardt, H. T. 1976. "Ideology and Etiology." *Journal of Medicine and Philosophy* 1 (3): 256–68. https://doi.org/10.1093/jmp/1.3.256.

Garson, Justin. 2016. *A Critical Overview of Biological Functions*.

———. 2019. *What Biological Functions Are and Why They Matter*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108560764.

Garson, Justin, and Gualtiero Piccinini. 2014. "Functions Must Be Performed at Appropriate Rates in Appropriate Situations." *The British Journal for the Philosophy of Science* 65 (1): 1–20. https://doi.org/10.1093/bjps/axs041.

Godfrey-Smith, Peter. 1993. "Functions: Consensus without Unity." *Pacific Philosophical Quarterly* 74 (3): 196–208. https://doi.org/10.1111/j.1468-0114.1993.tb00358.x.

Goosens, William K. 1980. "Values, Health, and Medicine." *Philosophy of Science* 47 (1): 100–115. https://doi.org/10.1086/288912.

Green, Sara. 2018. "Philosophy of Systems and Synthetic Biology." The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (Ed.). 2018. <https://plato.stanford.edu/archives/sum2018/entries/systems-synthetic-biology/>.

Griffiths, Paul E., and John Matthewson. 2016. "Evolution, Dysfunction, and Disease: A Reappraisal." *The British Journal for the Philosophy of Science* 69 (2): 301–27. https://doi.org/10.1093/bjps/axw021.

Hausman, D. M. 2014. "Health and Functional Efficiency." *Journal of Medicine and Philosophy* 39 (6): 634–47. https://doi.org/10.1093/jmp/jhu036.

Hauswald, Rico, and Lara Keuc. 2017. "Indeterminacy in Medical Classification: On Continuity, Uncertainty, and Vagueness." In *Vagueness in Psychiatry,* edited by Geert Keil, Lara Keuck, and Rico Hauswald, 93–117. Oxford: Oxford University Press.

Hofmann, Bjørn. 2021. "How to Draw the Line between Health and Disease? Start with Suffering." *Health Care Analysis* 29 (2): 127–43. https://doi.org/10.1007/s10728-021-00434-0.

Hucklenbroich, Peter. 2017. "Disease Entities and the Borderline between Health and Disease: Where Is the Place of Gradations?" In *Vagueness in Psychiatry*, edited by Geert Keil, Lara Keuck, and Rico Hauswald, 75–92. Oxford: Oxford University Press.

Jutel, Annemarie. 2011. "Classification, Disease, and Diagnosis." *Perspectives in Biology and Medicine* 54 (2): 189–205. https://doi.org/10.1353/pbm.2011.0015.

Kingma, Elselijn. 2010. "Paracetamol, Poison, and Polio: Why Boorse's Account of Function Fails to Distinguish Health and Disease." *The British Journal for the Philosophy of Science* 61 (2): 241–64. https://doi.org/10.1093/bjps/axp034.

Kraemer, Daniel M. 2013. "Statistical Theories of Functions and the Problem of Epidemic Disease." *Biology & Philosophy* 28 (3): 423–38. https://doi.org/10.1007/s10539-013-9365-3.

Lemoine, Maël. 2013. "Defining Disease beyond Conceptual Analysis: An Analysis of Conceptual Analysis in Philosophy of Medicine." *Theoretical Medicine and Bioethics* 34 (4): 309–25. https://doi.org/10.1007/s11017-013-9261-5.

Matthewson, John. 2020. "Does Proper Function Come in Degrees?" *Biology & Philosophy* 35 (4): 39. https://doi.org/10.1007/s10539-020-09758-y.

Matthewson, John, and Paul E. Griffiths. 2017. "Biological Criteria of Disease: Four Ways of Going Wrong." *The Journal of Medicine and Philosophy* 42 (4): 447–66. https://doi.org/10.1093/jmp/jhx004.

McLaughlin, Peter. 2009. "Functions and Norms." In *Functions in Biological and Artificial Worlds: Comparative Philosophical Perspectives*, edited by Ulrich Krohs and Peter Kroes, Springer, 93-102.

Melander, Peter. 1997. "Analyzing Functions: An Essay on a Fundamental Notion in Biology." Acta Universitatis Umensis 138. Stockholm: Almqvist & Wiksell International.

Millikan, Ruth Garrett. 1989. "In Defense of Proper Functions." *Philosophy of Science* 56 (2): 288–302. https://doi.org/10.1086/289488.

Neander, Karen. 1991. "The Teleological Notion of "Function"." *Australasian Journal of Philosophy* 69 (4): 454–68. https://doi.org/10.1080/00048409112344881.

———. 1995. "Misrepresenting & Malfunctioning." *Philosophical Studies* 79 (2): 109–41. https://doi.org/10.1007/BF00989706.

Nordenfelt, Lennart. 1995. *On the Nature of Health: An Action-Theoretic Approach*. Dordrecht: Springer.

Rogers, Wendy A., and Mary Jean Walker. 2017. "The Line-Drawing Problem in Disease Definition." *The Journal of Medicine and Philosophy* 42 (4): 405–23. https://doi.org/10.1093/jmp/jhx010.

Saborido, Cristian, Alvaro Moreno, María González-Moreno, and Juan Carlos Hernández Clemente. 2016. "Organizational Malfunctions and the Notions of Health and Disease." In *Naturalism in the Philosophy of Health*, edited by Élodie Giroux, 101–20. Cham: Springer. https://doi.org/10.1007/978-3-319-29091-1_7.

Schramme, Thomas. 2010. "Can We Define Mental Disorder by Using the Criterion of Mental Dysfunction?" *Theoretical Medicine and Bioethics* 31 (1): 35–47. https://doi.org/10.1007/s11017-010-9136-y.

———. 2019. *Theories of Health Justice: Just Enough Health*. London New York: Rowman & Littlefield International.

Schroeder, S. Andrew. 2013. "Rethinking Health: Healthy or Healthier Than?" *The British Journal for the Philosophy of Science* 64 (1): 131–59. https://doi.org/10.1093/bjps/axs006.

Schwartz, Peter H. 2007. "Defining Dysfunction: Natural Selection, Design, and Drawing a Line*." *Philosophy of Science* 74 (3): 364–85. https://doi.org/10.1086/521970.

———. 2017. "Progress in Defining Disease: Improved Approaches and Increased Impact." *The Journal of Medicine and Philosophy* 42 (4): 485–502. https://doi.org/10.1093/jmp/jhx012.

Smith, John Maynard. 1995. "Genes, Memes, & Minds." *New York Review of Books*, 1995. https://www.nybooks.com/articles/1995/11/30/genes-memes-minds/.

Tresker, Steven. 2020. "Theoretical and Clinical Disease and the Biostatistical Theory." *Studies in History and Philosophy of*

Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 82: 101249. https://doi.org/10.1016/j.shpsc.2019.101249.

Varga, Somogy. 2020. "Epistemic Authority, Philosophical Explication, and the Bio-Statistical Theory of Disease." *Erkenntnis* 85 (4): 937–56. https://doi.org/10.1007/s10670-018-0058-9.

Veit, Walter. 2021. "Biological Normativity: A New Hope for Naturalism?" *Medicine, Health Care and Philosophy* 24 (2): 291–301. https://doi.org/10.1007/s11019-020-09993-w.

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): 648–82. https://doi.org/10.1093/jmp/jhu038.

Wakefield, Jerome C. 1992. "The Concept of Mental Disorder: On the Boundary between Biological Facts and Social Values." *American Psychologist* 47 (3): 373–88. https://doi.org/10.1037/0003-066X.47.3.373.

Whitbeck, Caroline. 1981. "What Is Diagnosis? Some Critical Reflections." *Metamedicine* 2 (3): 319–29. https://doi.org/10.1007/BF00882078.

Wouters, Arno G. 2003. "Four Notions of Biological Function." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 34 (4): 633–68. https://doi.org/10.1016/j.shpsc.2003.09.006.

Wright, Larry. 1973. "Functions." *The Philosophical Review* 82 (2), 139–68.

ABSTRACTS (SAŽECI)

FAMINE, AFFLUENCE, AND AMORALITY

David Sackris
Arapahoe Community College

ABSTRACT

I argue that the debate concerning the nature of first-person moral judgment, namely, whether such moral judgments are inherently motivating (internalism) or whether moral judgments can be made in the absence of motivation (externalism), may be founded on a faulty assumption: that moral judgments form a distinct kind that must have some shared, essential features in regards to motivation to act. I argue that there is little reason to suppose that first-person moral judgments form a homogenous class in this respect by considering an ordinary case: student readers of Peter Singer's "Famine, Affluence, and Morality". Neither internalists nor externalists can provide a satisfying account as to why our students fail to act in this particular case, but are motivated to act by their moral judgments in most cases. I argue that the inability to provide a satisfying account is rooted in this shared assumption about the nature of moral judgments. Once we consider rejecting the notion that first-person moral decision- making forms a distinct kind in the way it is typically assumed, the internalist/externalist debate may be rendered moot.

Keywords: meta-ethics; moral judgment; internalism; externalism; natural kinds

GLAD, BOGATSTVO I AMORALNOST

David Sackris
Arapahoe Community College

SAŽETAK

Tvrdim da se rasprava o prirodi moralnog prosuđivanja u prvom licu, preciznije, pitanja o tome jesu li takvi moralni sudovi inherentno motivirajući (internalizam) ili se moralni sudovi mogu donijeti u nedostatku motivacije (eksternalizam) mogu temeljiti na pogrešnoj pretpostavci: da moralni sudovi čine posebnu vrstu koja mora imati neke zajedničke, bitne značajke u pogledu motivacije za djelovanje. Tvrdim da nema razloga za pretpostavku da moralni sudovi iz prvog lica čine homogenu klasu razmatrajući običan slučaj: studenti koji čitaju "Famine,

Affluence, and Morality" Petera Singera. Ni internalisti ni eksternalisti ne mogu dati zadovoljavajuće objašnjenje zašto naši studenti ne postupaju u skladu sa svojim moralnim sudovima u ovom konkretnom slučaju, iako su u većini slučajeva motivirani djelovati u skladu sa svojim moralnim sudovima. Tvrdim da nemogućnost pružanja zadovoljavajućeg objašnjenja ima svoj izvor u uobičajenoj pretpostavci o prirodi moralnih sudova. Nakon što razmotrimo mogućnost odbacivanja tvrdnje da moralno odlučivanje iz prvog lica čini posebnu vrstu na način na koji se to obično pretpostavlja, rasprava o internalizmu/eksternalizmu se može smatrati spornom.

Ključne riječi: metaetika, moralni sud, internalizam, eksternalizam, prirodne vrste

LOGICAL RELATIVISM THROUGH LOGICAL CONTEXTS

Jonas R. Becker Arenhart
Federal University of Santa Catarina and Federal University of Maranhão

ABSTRACT

We advance an approach to logical contexts that grounds the claim that logic is a local matter: distinct contexts require distinct logics. The approach results from a concern about context individuation, and holds that a logic may be constitutive of a context or domain of application. We add a naturalistic component: distinct domains are more than mere technical curiosities; as intuitionistic mathematics testifies, some of the distinct forms of inference in different domains are actively pursued as legitimate fields of research in current mathematics, so, unless one is willing to revise the current scientific practice, generalism must go. The approach is advanced by discussing some tenets of a similar argument advanced by Shapiro, in the context of logic as models approach. In order to make our view more appealing, we reformulate a version of logic as models approach following naturalistic lines, and bring logic closer to the use of models in science.

Keywords: classical logic; intuitionistic logic; relativism; logic as models; context constitution

LOGIČKI RELATIVIZAM KROZ LOGIČKE KONTEKSTE

Jonas R. Becker Arenhart
Federal University of Santa Catarina and Federal University of Maranhão

SAŽETAK

Unaprjeđujemo pristup logičkim kontekstima koji utemeljuju tvrdnju da je logika lokalna stvar: različiti konteksti zahtijevaju različite logike. Pristup proizlazi iz brige oko individuacije konteksta i smatra da logika može biti konstitutivna za kontekst ili domenu primjene. Dodajemo naturalističku komponentu: različite domene su više od pukih tehničkih zanimljivosti. Kao što intuicionistička matematika svjedoči, neki od različitih oblika zaključivanja u različitim domenama, aktivno se slijede kao legitimna polja istraživanja u aktualnoj matematici, stoga, osim ako netko nije voljan revidirati trenutnu znanstvenu praksu, generalizam se mora napustiti. Pristup je unaprijeđen raspravom o nekim načelima sličnog argumenta koje je iznio Shapiro, u kontekstu pristupa logike kao modela. Kako bismo naš pristup učinili privlačnijim, preformulirali smo verziju logike kao pristup modela, slijedeći naturalističke linije i približili logiku korištenju modela u znanosti.

Ključne riječi: klasična logika; intuicionistička logika; relativizam; logika kao modeli; konstitutivni konteksti

INTRODUCTION TO THE BOOK SYMPOSIUM ON THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE BY GUEST EDITORS

Maria Cristina Amoretti
University of Genoa

Elisabetta Lalumera
University of Bologna

ABSTRACT

Introduction to the book symposium “THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE: NEW PHILOSOPHICAL AND SCIENTIFIC DEVELOPMENTS BY DEREK BOLTON AND GRANT GILLETT”.

Keywords: Biopsychosocial model, medical disorder, Derek Bolton, Grant Gillett

UVOD GOSTUJUĆIH UREDNIKA U SIMPOZIJ O BIOPSIHOSOCIJALNOM MODELU ZDRAVLJA I BOLESTI

Maria Cristina Amoretti
University of Genoa

Elisabetta Lalumera
University of Bologna

SAŽETAK

Uvod u simpozij o knjizi „Biopsihosocijalni model zdravlja i bolesti: novi filozofski i znanstveni razvoj, Derek Bolton i Grant Gillett".

Ključne riječi: Biopsihosocijalni model, medicinski poremećaj, Derek Bolton, Grant Gillett

FROM ENGEL TO ENACTIVISM: CONTEXTUALIZING THE BIOPSYCHOSOCIAL MODEL

Awais Aftab
Case Western Reserve University

Kristopher Nielsen
Victoria University of Wellington

ABSTRACT

In this article we offer a two-part commentary on Bolton and Gillett's reconceptualization of Engel's biopsychosocial model. In the first section we present a conceptual and historical assessment of the biopsychosocial model that differs from the analysis by Bolton and Gillett. Specifically, we point out that Engel in his vision of the biopsychosocial model was less concerned with the ontological possibility and nature of psychosocial causes, and more concerned with psychosocial influences in the form of illness interpretation and presentation, sick role, seeking or rejection of care, the doctor-patient therapeutic relationship, and role of personality factors and family relationships in recovery from illness, etc. On the basis of this assessment, we then question Bolton and Gillett's restricted focus on accounting for biopsychosocial causal interactions. The second section compares Bolton and Gillett's account with a recent enactivist account of mental disorder that tackles similar conceptual problems of causal interactions. Bolton and Gillett's utilize elements of the 4E cognition, but they combine these proto-ideas with an information-processing paradigm.

Given their explicit endorsement of 4E approaches to mind and cognition, we illustrate some key ways in which a more fleshed out enactive account, particularly one that doesn't rely on notions of information-processing, differs from the account proposed by Bolton and Gillett.

Keywords: biopsychosocial model; George Engel; causality; enactivism; 4E cognition

OD ENGELA DO ENAKTIVIZMA: KONTEKSTUALIZACIJA BIOPSIHOSOCIJALNOG MODELA

Awais Aftab
Case Western Reserve University

Kristopher Nielsen
Victoria University of Wellington

SAŽETAK

U ovom članku nudimo dvodijelni komentar na Boltonovu i Gillettovu rekonceptualizaciju Engelovog biopsihosocijalnog modela. U prvom dijelu predstavljamo pojmovnu i povijesnu procjenu biopsihosocijalnog modela koja se razlikuje od Boltonove i Gillettove analize. Konkretnije, ističemo da se Engel u svojoj viziji biopsihosocijalnog modela manje bavio ontološkom mogućnošću i prirodom psihosocijalnih uzroka, a više se bavio psihosocijalnim utjecajima u obliku interpretacije i prezentacije bolesti, uloge bolesnika, traženja ili odbijanja skrbi, terapijski odnos liječnik-pacijent, te uloga osobnosti i obiteljskih odnosa u oporavku od bolesti, itd. Na temelju ove procjene onda dovodimo u pitanje ograničeni fokus Boltona i Gilletta na objašnjenje biopsihosocijalnih uzročno-posljedičnih interakcija. Drugi dio uspoređuje Boltonovu i Gillettovu teoriju s nedavnim enaktivističkom teorijom mentalnog poremećaja koja se bavi sličnim pojmovnim problemima uzročno-posljedičnih interakcija. Bolton i Gillett koriste elemente 4E spoznaje, ali kombiniraju ove proto-ideje s paradigmom obrade informacija. S obzirom na njihovo eksplicitno prihvaćanje 4E pristupa umu i spoznaji, ilustriramo neke ključne načine na koje se detaljniji enaktivno objašnjenje, osobito ono koje se ne oslanja na pojmove obrade informacija, razlikuje od objašnjenja koje su predložili Bolton i Gillett.

Ključne riječi: biopsihosocijalni model; George Engel; uzročnost; enaktivizam; 4E spoznaja

CENTRIFUGAL AND CENTRIPETAL THINKING ABOUT THE BIOPSYCHOSOCIAL MODEL IN PSYCHIATRY

Kathryn Tabb
Philosophy Program, Bard College

ABSTRACT

The biopsychosocial model, which was deeply influential on psychiatry following its introduction by George L. Engel in 1977, has recently made a comeback. Derek Bolton and Grant Gillett have argued that Engel's original formulation offered a promising general framework for thinking about health and disease, but that this promise requires new empirical and philosophical tools in order to be realized. In particular, Bolton and Gillett offer an original analysis of the ontological relations between Engel's biological, social, and psychological levels of analysis. I argue that Bolton and Gillett's updated model, while providing an intriguing new metaphysical framework for medicine, cannot resolve some of the most vexing problems facing psychiatry, which have to do with how to prioritize different sorts of research. These problems are fundamentally ethical, rather than ontological. Without the right prudential motivation, in other words, the unification of psychiatry under a single conceptual framework seems doubtful, no matter how compelling the model. An updated biopsychosocial model should include explicit normative commitments about the aims of medicine that can give guidance about the sorts of causal connections to be prioritized as research and clinical targets.

Keywords: Biopsychosocial model; precision medicine, medical ethics; philosophy of psychiatry

CENTRIFUGALNO I CENTRIPETALNO RAZMIŠLJANJE O BIOPSIHOSOCIJALNOM MODELU U PSIHIJATRIJI

Kathryn Tabb
Philosophy Program, Bard College

SAŽETAK

Biopsihosocijalni model, koji je imao dubok utjecaj na psihijatriju nakon što ga je uveo George L. Engel 1977, nedavno se vratio. Derek Bolton i Grant Gillett tvrde da je Engelova izvorna formulacija ponudila obećavajući opći okvir za razmišljanje o zdravlju i bolesti, ali da to obećanje zahtijeva nove empirijske i filozofske alate kako bi se ostvarilo.

Bolton i Gillett nude originalnu analizu ontoloških odnosa između Engelove biološke, društvene i psihološke razine analize. Argumentiram da Boltonov i Gillettov ažurirani model, iako pruža intrigantan novi metafizički okvir za medicinu, ne može riješiti neke od najzahtjevnijih problema s kojima se psihijatrija suočava, a koji se odnose na to kako dati prioritet različitim vrstama istraživanja. Ti su problemi u osnovi etički, a ne ontološki. Bez prave prudencijalne motivacije, drugim riječima, objedinjavanje psihijatrije pod jednim pojmovnim okvirom čini se upitnim, ma koliko uvjerljiv model. Ažurirani biopsihosocijalni model trebao bi uključivati eksplicitne normativne obveze o ciljevima medicine koji mogu dati smjernice o vrstama uzročno-posljedičnih veza kojima se treba dati prioritet kao istraživačkim i kliničkim ciljevima.

Ključne riječi: biopsihosocijalni model; precizna medicina; medicinska etika; filozofija psihijatrije

HOW TO BE A HOLIST WHO REJECTS THE BIOPSYCHOSOCIAL MODEL

Diane O'Leary
Center for Philosophy of Science, University of Pittsburgh

ABSTRACT

After nearly fifty years of mea culpas and explanatory additions, the biopsychosocial model is no closer to a life of its own. Bolton and Gillett give it a strong philosophical boost in The Biopsychosocial Model of Health and Disease, but they overlook the model's deeply inconsistent position on dualism. Moreover, because metaphysical confusion has clinical ramifications in medicine, their solution sidesteps the model's most pressing clinical faults. But the news is not all bad. We can maintain the merits of holism as we let go of the inchoate bag of platitudes that is the biopsychosocial model. We can accept holism as the metaphysical open door that it is, just a willingness to recognize the reality of human experience, and the sense in which that reality forces medicine to address biological, psychological, and social aspects of health. This allows us to finally characterize Engel's driving idea in accurate philosophical terms, as acceptance of (phenomenal) consciousness in the context of medical science. This will not entirely pin down medicine's stance on dualism, but it will position it clearly enough to readily improve patient care.

Keywords: Biopsychosocial model; holism; dualism; philosophy of medicine; psychosomatic medicine

KAKO BITI HOLIST KOJI ODBACUJE BIOPSIHOSOCIJALNI MODEL

Diane O'Leary
Center for Philosophy of Science, University of Pittsburgh

SAŽETAK

Nakon gotovo pedeset godina mea culpa i objašnjavajućih dodataka, biopsihosocijalni model nije ništa bliži vlastitom životu. Bolton i Gillett daju mu snažan filozofski poticaj u „Biopsihosocijalnom modelu zdravlja i bolesti", ali zanemaruju duboko nedosljedan stav koji model ima prema dualizmu. Štoviše, budući da metafizička zbrka ima kliničke posljedice u medicini, njihovo rješenje zaobilazi najhitnije kliničke greške modela. Međutim, nije sve tako crno. Možemo zadržati dobre strane holizma istovremeno napuštajući floskule koje pretpostavlja biopsihosocijalni model. Možemo prihvatiti holizam kao metafizički otvorena vrata koja on jest, samo spremnost da se prepozna stvarnost ljudskog iskustva i smisao u kojem ta stvarnost tjera medicinu da se pozabavi biološkim, psihološkim i društvenim aspektima zdravlja. To nam omogućuje da konačno okarakteriziramo Engelovu pokretačku ideju u točnim filozofskim terminima, kao prihvaćanje (fenomenalne) svijesti u kontekstu medicinske znanosti. To neće u potpunosti odrediti stav medicine prema dualizmu, ali će ga postaviti dovoljno jasno da se lako poboljša skrb za pacijente.

Ključne riječi: biopsihosocijalni model; holizam; dualizam; filozofija medicine; psihosomatska medicina

CAUSATION AND CAUSAL SELECTION IN THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE

Hane Htut Maung
University of Manchester

ABSTRACT

In The Biopsychosocial Model of Health and Disease, Derek Bolton and Grant Gillett argue that a defensible updated version of the biopsychosocial model requires a metaphysically adequate account of disease causation that can accommodate biological, psychological, and social factors. This present paper offers a philosophical critique of their account of biopsychosocial causation. I argue that their account relies on claims about the normativity and the semantic content of biological information that are

metaphysically contentious. Moreover, I suggest that these claims are unnecessary for a defence of biopsychosocial causation, as the roles of multiple and diverse factors in disease causation can be readily accommodated by a more widely accepted and less metaphysically contentious account of causation. I then raise the more general concern that they are misdiagnosing the problem with the traditional version of the biopsychosocial model. The challenge when developing an explanatorily valuable version of the biopsychosocial model, I argue, is not so much providing an adequate account of biopsychosocial causation, but providing an adequate account of causal selection. Finally, I consider how this problem may be solved to arrive at a more explanatorily valuable and clinically useful version of the biopsychosocial model.

Keywords: Derek Bolton; Grant Gillett; biopsychosocial model; causation; causal selection

UZROČNOST I UZROČNA SELEKCIJA U BIOPSIHOSOCIJALNOM MODELU ZDRAVLJA I BOLESTI

Hane Htut Maung
University of Manchester

SAŽETAK

U „The Biopsychosocial Model of Health and Disease“, Derek Bolton i Grant Gillett tvrde da obranjiva ažurirana verzija biopsihosocijalnog modela zahtijeva metafizički adekvatnu teoriju uzroka bolesti koja može zahvatiti biološke, psihološke i socijalne čimbenike. Ovaj rad nudi filozofsku kritiku njihove teorije biopsihosocijalne uzročnosti. Tvrdim da se njihova teorija oslanja na tvrdnje o normativnosti i semantičkom sadržaju bioloških informacija koje su metafizički sporne. Štoviše, sugeriram da su ove tvrdnje nepotrebne za obranu biopsihosocijalne uzročnosti, budući da se uloge višestrukih i raznolikih čimbenika u uzrokovanju bolesti mogu lako prilagoditi naširoko prihvaćenom i manje metafizički spornom teorijom uzročnosti. Zatim iznosim općenitiji prigovor da Bolton i Gillett pogrešno dijagnosticiraju problem s tradicionalnom verzijom biopsihosocijalnog modela. Tvrdim da izazov pri razvijanju eksplanatorno vrijedne verzije biopsihosocijalnog modela nije toliko pružanje adekvatne teorije biopsihosocijalne uzročnosti, već pružanje odgovarajuće teorije uzročne selekcije. Konačno, razmatram kako se ovaj problem može riješiti kako bismo došli do eksplanatorno vrjednije i klinički korisnije verzije biopsihosocijalnog modela.

Ključne riječi: Derek Bolton; Grant Gillett; biopsihosocijalni model; uzročnost; uzročni selekcija

THE BIOPSYCHOSOCIAL MODEL OF HEALTH AND DISEASE: RESPONSES TO THE 4 COMMENTARIES

Derek Bolton
King's College London

ABSTRACT

I respond to the 4 commentaries by Awais Aftab & Kristopher Nielsen (A&N), Hane Htut Maung (HHM), Diane O'Leary (DO'L) and Kathryn Tabb (KT) under 3 main headings: "What is the BPSM really?" & Why update it?; "Is our approach foundationally compromised?", and finally, "Antagonists or fellow travellers?".

Keywords: Biopsychosocial model; causation; George Engel; information

BIOPSIHOSOCIJALNI MODEL ZDRAVLJA I BOLESTI: ODGOVORI NA 4 KOMENTARA

Derek Bolton
King's College London

SAŽETAK

Odgovaram na komentare Awaisa Aftaba i Kristophera Nielsena (A&N), Hane Htut Maunga (HHM), Diane O'Leary (DO'L) i Kathryn Tabb (KT) pod trima glavna naslova: „Što je zapravo BPSM? ", „Zašto ga ažurirati?", „Je li naš pristup temeljno ugrožen?" i posljednje, „Antagonisti ili suputnici?".

INTRODUCTION TO THE SPECIAL ISSUE ON PHILOSOPHY OF MEDICINE

Saana Jukola
University of Bonn

Anke Bueter
Aarhus University

ABSTRACT

This article is an introduction to the special issue on philosophy of medicine. Philosophy of medicine is a field that has flourished in the last couple of decades and has become increasingly institutionalized. The introduction begins with a brief overview of some of the most central recent developments in the field. It then describes the six articles that comprise this issue.

Keywords: philosophy of medicine; medical ethics; medical epistemology; disease; diagnosis

UVOD U POSEBNO IZDANJE O FILOZOFIJI MEDICINE

Saana Jukola
University of Bonn

Anke Bueter
Aarhus University

SAŽETAK

Ovaj je članak uvod u posebno izdanje o filozofiji medicine. Filozofija medicine je područje koje je procvjetalo u posljednjih nekoliko desetljeća i postaje sve više institucionalizirano. Uvod započinje kratkim pregledom nekih od najvažnijih nedavnih razvoja na tom području. Zatim se opisuju šest članaka koji obuhvaćaju ovo pitanje.

Ključne riječi: filozofija medicine, medicinska etika, medicinska epistemologija, bolest, dijagnoza

DIAGNOSTIC JUSTICE: TESTING FOR COVID-19

Ashley Graham Kennedy
Florida Atlantic University

Bryan Cwik
Portland State University

ABSTRACT

Diagnostic testing can be used for many purposes, including testing to facilitate the clinical care of individual patients, testing as an inclusion

criterion for clinical trial participation, and both passive and active surveillance testing of the general population in order to facilitate public health outcomes, such as the containment or mitigation of an infectious disease. As such, diagnostic testing presents us with ethical questions that are, in part, already addressed in the literature on clinical care as well as clinical research (such as the rights of patients to refuse testing or treatment in the clinical setting or the rights of participants in randomized controlled trials to withdraw from the trial at any time). However, diagnostic testing, for the purpose of disease surveillance also raises ethical issues that we do not encounter in these settings, and thus have not been much discussed. In this paper we will be concerned with the similarities and differences between the ethical considerations in these three domains: clinical care, clinical research, and public health, as they relate to diagnostic testing specifically. Via an examination of the COVID-19 case we will show how an appeal to the concept of diagnostic justice helps us to make sense of the (at times competing) ethical considerations in these three domains.

Keywords: diagnostic justice; philosophy of medicine; political philosophy; applied ethics

DIJAGNOSTIČKA PRAVDA: TESTIRANJE NA COVID-19

Ashley Graham Kennedy
Florida Atlantic University

Bryan Cwik
Portland State University

SAŽETAK

Dijagnostičko testiranje može se koristiti u mnoge svrhe, uključujući testiranje za olakšavanje kliničke skrbi pojedinačnih pacijenata, testiranje kao kriterij uključivanja za sudjelovanje u kliničkim ispitivanjima te kao pasivno i aktivno nadzorno testiranje opće populacije kako bi se olakšali ishodi javnog zdravlja, kao što su obuzdavanje ili ublažavanje zarazne bolesti. Kao takvo, dijagnostičko testiranje nam postavlja etička pitanja koja su dijelom već obrađena u literaturi o kliničkoj skrbi, kao i kliničkim istraživanjima (kao što su prava pacijenata da odbiju testiranje ili liječenje u kliničkom okruženju ili prava sudionika u nasumičnim, kontroliranim ispitivanjima da se povuku iz ispitivanja u bilo kojem trenutku). Međutim, dijagnostičko testiranje, u svrhu nadzora bolesti, postavlja i etička pitanja s kojima se ne susrećemo u ovim okruženjima, pa se o njima nije puno raspravljalo. U ovom radu bavit ćemo se sličnostima i razlikama između etičkih razmatranja u tri domene: kliničkoj skrbi, kliničkim istraživanjima

i javnom zdravstvu jer se one posebno odnose na dijagnostičko testiranje. Kroz ispitivanje slučaja COVID-19 pokazat ćemo kako nam pozivanje na pojam dijagnostičke pravde pomaže da shvatimo (ponekad suparnička) etička razmatranja u ovim trima domenama.

Ključne riječi: dijagnostička pravda, filozofija medicine, politička filozofija, primijenjena etika

ADAPT TO TRANSLATE – ADAPTIVE CLINICAL TRIALS AND BIOMEDICAL INNOVATION

Daria Jadreškić
University of Klagenfurt

ABSTRACT

The article presents the advantages and limitations of adaptive clinical trials for assessing the effectiveness of medical interventions and specifies the conditions that contributed to their development and implementation in clinical practice. I advance two arguments by discussing different cases of adaptive trials. The normative argument is that responsible adaptation should be taken seriously as a new way of doing clinical research insofar as a valid justification, sufficient understanding, and adequate operational conditions are provided. The second argument is historical. The development of adaptive trials can be related to lessons learned from research in cases of urgency and to the decades-long efforts to end the productivity crisis of pharmaceutical research, which led to the emergence of translational, personalized, and, recently, precision medicine movements.

Keywords: adaptive clinical trials; randomized controlled trials; reliability; urgency; precision medicine; translational medicine; the productivity crisis

PRILAGODBA ZA TRANSLACIJU – PRILAGODLJIVA KLINIČKA ISPITIVANJA I BIOMEDICINSKE INOVACIJE

Daria Jadreškić
University of Klagenfurt

SAŽETAK

U članku su prikazane prednosti i ograničenja adaptivnih kliničkih ispitivanja za procjenu učinkovitosti medicinskih intervencija te se specificiraju uvjeti koji su pridonijeli njihovom razvoju i primjeni u kliničkoj praksi. Iznosim dva argumenta na temelju rasprave različitih slučajeva adaptivnih ispitivanja. Normativni argument je da se odgovornu prilagodbu treba shvatiti ozbiljno kao novi način kliničkog istraživanja u mjeri u kojoj je osigurano valjano opravdanje, dovoljno razumijevanja i odgovarajući operativni uvjeti. Drugi argument je povijesni. Razvoj adaptivnih ispitivanja može se povezati s lekcijama naučenima iz istraživanja u slučajevima hitnosti i desetljećima dugim naporima da se okonča kriza produktivnosti farmaceutskih istraživanja, koja je dovela do pojave translacijskih, personaliziranih i, nedavno, pokreta precizne medicine.

Ključne riječi: adaptivna klinička ispitivanja, nasumična kontrolirana ispitivanja, pouzdanost, hitnost, precizna medicina, translacijska medicina, kriza produktivnosti

WRONGFUL MEDICALIZATION AND EPISTEMIC INJUSTICE IN PSYCHIATRY: THE CASE OF PREMENSTRUAL DYSPHORIC DISORDER

Anne-Marie Gagné-Julien
Biomedical Ethics Unit, McGill University

ABSTRACT

In this paper, my goal is to use an epistemic injustice framework to extend an existing normative analysis of over-medicalization to psychiatry and thus draw attention to overlooked injustices. Kaczmarek (2019) has developed a promising bioethical and pragmatic approach to over-medicalization, which consists of four guiding questions covering issues related to the harms and benefits of medicalization. In a nutshell, if we answer "yes" to all proposed questions, then it is a case of over-medicalization. Building on an epistemic injustice framework, I will argue that Kaczmarek's proposal lacks guidance concerning the procedures through which we are to answer the four questions, and I will import the conceptual resources of epistemic injustice to guide our thinking on these issues. This will lead me to defend more inclusive decision-making procedures regarding medicalization in the DSM. Kaczmarek's account complemented with an epistemic injustice framework can help us achieve

better forms of medicalization. I will then use a contested case of medicalization, the creation of Premenstrual Dysphoric Disorder (PMDD) in the DSM-5 to illustrate how the epistemic injustice framework can help to shed light on these issues and to show its relevance to distinguish good and bad forms of medicalization.

Keywords: over-medicalization; epistemic injustice; premenstrual dysphoric disorder; hermeneutical injustice; pre-emptive testimonial injustice; Miranda Fricker

POGREŠNA MEDIKALIZACIJA I EPISTEMIČKA NEPRAVDA U PSIHIJATRIJI: SLUČAJ PREDMENSTRUALNOG DISFORIČNOG POREMEĆAJA

Anne-Marie Gagné-Julien
Biomedical Ethics Unit, McGill University

SAŽETAK

U ovom radu, cilj mi je upotrijebiti okvir epistemički nepravde kako bih proširila postojeću normativnu analizu pretjerane medikalizacije na psihijatriju i tako skrenula pozornost na zanemarene nepravde. Kaczmarek (2019) razvija obećavajući bioetički i pragmatičan pristup pretjeranoj medikalizaciji, koji se sastoji od četiri pitanja koja pokrivaju probleme vezane za štete i prednosti medikalizacije. Ukratko, ako na sva predložena pitanja odgovorimo s "da", onda je riječ o pretjeranoj medikalizaciji. Nadovezujući se na okvir epistemičke nepravde, tvrdit ću da Kaczmarekovom prijedlogu nedostaju smjernice u vezi s postupcima kojima trebamo odgovoriti na četiri pitanja i uvest ću pojmovne resurse epistemičke nepravde kako bi usmjerili naše razmišljanje o tim pitanjima. To će me navesti da branim inkluzivnije postupke donošenja odluka u vezi s medikalizacijom u DSM-u. Kaczmarekovo gledište dopunjeno okvirom epistemičke nepravde može nam pomoći da postignemo bolje oblike medikalizacije. Zatim ću upotrijebiti sporni slučaj medikalizacije, stvaranje predmenstrualnog disforičnog poremećaja (PMDD) u DSM-5, kako bih ilustrirala na koji način okvir epistemičke nepravde može pomoći u rasvjetljavanju ovih problema i pokazati njegovu relevantnost za razlikovanje dobrih i loših oblika medikalizacije.

Ključne riječi: prekomjerna medikalizacija, epistemička nepravda, predmenstrualni disforični poremećaj, hermeneutička nepravda, preventivna svjedočanska nepravda, Miranda Fricker

MEDICALIZATION OF SEXUAL DESIRE

Jacob Stegenga
University of Cambridge

ABSTRACT

Medicalisation is a social phenomenon in which conditions that were once under legal, religious, personal or other jurisdictions are brought into the domain of medical authority. Low sexual desire in females has been medicalised, pathologised as a disease, and intervened upon with a range of pharmaceuticals. There are two polarised positions on the medicalisation of low female sexual desire: I call these the mainstream view and the critical view. I assess the central arguments for both positions. Dividing the two positions are opposing models of the aetiology of low female sexual desire. I conclude by suggesting that the balance of arguments supports a modest defence of the critical view regarding the medicalisation of low female sexual desire.

Keywords: medicalization; female sexual interest/arousal disorder; philosophy of medicine; disease; controversial diseases; philosophy of psychiatry

MEDIKALIZACIJA SEKSUALNE ŽELJE

Jacob Stegenga
University of Cambridge

SAŽETAK

Medikalizacija je društveni fenomen u kojem se uvjeti koji su nekada bili pod zakonskom, vjerskom, osobnom ili drugom jurisdikcijom stavljaju u domenu medicinskog autoriteta. Niska seksualna želja kod žena je medikalizirana, patologizirana kao bolest te se tretira nizom lijekova. Postoje dvije polarizirane pozicije o medikalizaciji niske ženske seksualne želje, nazivam ih: mainstream gledište i kritičko gledište. Ocjenjujem središnje argumente za obje pozicije. Ono što dijeli ove dvije pozicije su suprotstavljeni modeli etiologije niske ženske seksualne želje. Zaključujem sugestijom da ravnoteža argumenata podržava skromnu obranu kritičkog stajališta o medikalizaciji niske ženske seksualne želje.

Ključne riječi: medikalizacija, ženski spolni interes/poremećaj uzbuđenja, filozofija medicine, bolest, kontroverzne bolesti, filozofija psihijatrije

WHEN A HYBRID ACCOUNT OF DISORDER IS NOT ENOUGH: THE CASE OF GENDER DYSPHORIA

Kathleen Murphy-Hollies
University of Birmingham

ABSTRACT

In this paper I discuss Wakefield's account of mental disorder as applied to the case of gender dysphoria (GD). I argue that despite being a hybrid account which brings together a naturalistic and normative element in order to avoid pathologising normal or expectable states, the theory alone is still not extensive enough to answer the question of whether GD should be classed as a disorder. I suggest that the hybrid account falls short in adequately investigating how the harm and dysfunction in cases of GD relate to each other, and secondly that the question of why some dysfunction is disvalued and experienced as harmful requires further consideration. This masks further analysis of patients' distress and results in an unhelpful overlap of two types of clinical patients within a diagnosis of GD; those with gender-role dysphoria and those with sex dysphoria. These two conditions can be associated with different harms and dysfunctions but Wakefield's hybrid account does not have the tools to recognise this. This misunderstanding of the sources of dysfunction and harm in those diagnosed with GD risks ineffective treatment for patients and reinforcing the very same prejudiced norms which were conducive to the state being experienced as harmful in the first place. The theory needs to engage, to a surprising and so far unacknowledged extent, with sociological concepts such as the categorisation and stratification of groups in society and the mechanism of systemic oppression, in order to answer the question of whether GD should be classed as a mental disorder. Only then can it successfully avoid pathologising normal or expectable states, as has been seen in past 'illnesses' such as homosexuality and 'drapetomania'.

Keywords: mental disorder; Wakefield; hybrid; gender dysphoria; DSM

KADA HIBRIDNA TEORIJA POREMEĆAJA NIJE DOVOLJNA: SLUČAJ RODNE DISFORIJE

Kathleen Murphy-Hollies
University of Birmingham

SAŽETAK

U ovom radu raspravljam o Wakefieldovoj teoriji mentalnog poremećaja primijenjenoj na slučaj rodne disforije (RD). Tvrdim da sama teorija, unatoč tome što je hibridna teorija koja povezuje naturalistički i normativni element kako bi se izbjegla patološka pojava normalnih ili očekivanih stanja, još uvijek nije dovoljno opsežna da odgovori na pitanje treba li RD klasificirati kao poremećaj. Sugeriram da hibridna teorija ne uspijeva na adekvatan način istražiti kako su šteta i disfunkcija u slučajevima RD međusobno povezane, a drugo da pitanje zašto je neka disfunkcija nepoželjna te se doživljava kao štetna zahtijeva daljnje razmatranje. To zamagljuje daljnju analizu patnje pacijenata i rezultira beskorisnim preklapanjem dviju vrsta kliničkih pacijenata unutar dijagnoze RD: oni s disforijom rodnih uloga i oni sa spolnom disforijom. Ova dva stanja mogu biti povezana s različitim štetama i disfunkcijama, ali Wakefieldova hibridna teorija nema sredstva za to prepoznati. Ovo nerazumijevanje izvora disfunkcije i štete kod onih s dijagnozom RD riskira neučinkovito liječenje pacijenata i jačanje istih normi predrasuda koje su dovele do toga da se stanje uopće doživljava kao štetno. Teorija se u iznenađujućoj i dosad nepriznatoj mjeri treba baviti sociološkim pojmovima kao što su kategorizacija i stratifikacija grupa u društvu i mehanizam sistemske opresije, kako bi se odgovorilo na pitanje treba li RD svrstati u mentalne poremećaje. Tek tada može uspješno izbjeći patologiziranje normalnih ili očekivanih stanja, kao što je viđeno u povijesnim slučajevima 'bolesti' kao što su homoseksualnost i 'drapetomanija'.

Ključne riječi: mentalni poremećaj, Wakefield, hibrid, spolna disforija, DSM

THE QUANTITATIVE PROBLEM FOR THEORIES OF DYSFUNCTION AND DISEASE

Thomas Schramme
University of Liverpool

ABSTRACT

Many biological functions allow for grades. For example, secretion of a specific hormone in an organism can be on a higher or lower level, compared to the same organism at another occasion or compared to other organisms. What levels of functioning constitute instances of dysfunction; where should we draw the line? This is the quantitative problem for theories of dysfunction and disease. I aim to defend a version of biological

theories of dysfunction to tackle this problem. However, I will also allow evaluative considerations to enter into a theory of disease. My argument is based on a distinction between a biological and a clinical perspective. Disease, according to my reasoning, is restricted to instances that fall within the boundaries of biological dysfunctions. Responding to the quantitative problem does not require arbitrary decisions or social value-judgements. Hence, I argue for a non-arbitrary, fact-based method to address the quantitative problem. Still, not all biological dysfunctions are instances of disease. Adding a clinical perspective allows us to prevent the potential over-inclusiveness of the biological perspective, because it restricts the boundaries of disease even further.

Keywords: theory of function; dysfunction; line-drawing problem; concept of disease; nosology

KVANTITATIVNI PROBLEM ZA TEORIJE DISFUNKCIJE I BOLESTI

Thomas Schramme
University of Liverpool

SAŽETAK

Mnoge biološke funkcije dopuštaju stupnjevanje. Na primjer, lučenje određenog hormona u organizmu može biti na višoj ili nižoj razini, u usporedbi s istim organizmom u drugim okolnostima ili u usporedbi s drugim organizmima. Koje razine funkcioniranja predstavljaju slučajeve disfunkcije: gdje povlačimo crtu? To je kvantitativni problem za teorije disfunkcije i bolesti. Cilj mi je braniti verziju bioloških teorija disfunkcije kako bih se uhvatio u koštac s ovim problemom. Međutim, također ću dopustiti da evaluativna razmatranja uđu u teoriju bolesti. Moj argument se temelji na razlikovanju između biološke i kliničke perspektive. Prema mom mišljenju, bolest je ograničena na slučajeve koji spadaju u granice bioloških disfunkcija. Odgovor na kvantitativni problem ne zahtijeva proizvoljne odluke ili društveno vrijednosne sudove. Stoga se zalažem za nearbitrarnu metodu koja se temelji na činjenicama kako bi se riješio kvantitativni problem. Ipak, nisu sve biološke disfunkcije instance bolesti. Dodavanje kliničke perspektive omogućuje nam da spriječimo potencijalnu preveliku uključenost biološke perspektive, zato što postavlja dodatna ograničenja za određivanje granica bolesti.

Ključne riječi: teorija funkcije, disfunkcija, problem određivanja granica, pojam bolesti, nozologija

Translated by Marko Jurjako (Rijeka) and Iva Martinić (Rijeka)

Proofread by Iva Martinić (Rijeka)

AUTHOR GUIDELINES

Publication ethics

EuJAP subscribes to the publication principles and ethical guidelines of the Committee on Publication Ethics (COPE).

Submitted manuscripts ought to:

- be unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere;
- be approved by all co-Authors;
- contain citations and references to avoid plagiarism, self-plagiarism, and illegitimate duplication of texts, figures, etc. Moreover, Authors should obtain permission to use any third party images, figures and the like from the respective copyright holders. The pre-reviewing process includes screening for plagiarism and self-plagiarism by means of internet browsing and software Turnitin;
- be sent exclusively electronically to the Editors (eujap@ffri.uniri.hr) (or to the Guest editors in the case of a special issue) in a Word compatible format;
- be prepared for blind refereeing: authors' names and their institutional affiliations should not appear on the manuscript. Moreover, "identifiers" in MS Word Properties should be removed;
- be accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 300 words), keywords, academic affiliation and full address for correspondence including e-mail address, and, if needed, a disclosure of the Authors' potential conflict of interest that might affect the conclusions, interpretation, and evaluation of the relevant work under consideration;
- be in American or British English;
- be no longer than 9000 words, including references (for Original and Review Articles).
- be between 2000 and 5000 words, including footnotes and references (for Discussions and Critical notices)

Malpractice statement

If the manuscript does not match the scope and aims of EuJAP, the Editors reserve the right to reject the manuscript without sending it out to external reviewers. Moreover, the Editors reserve the right to reject submissions that do not satisfy any of the previous conditions.

If, due to the authors' failure to inform the Editors, already published material will appear in EuJAP, the Editors will report the authors' unethical behaviour in the next issue and remove the publication from EuJAP web site and the repository HRČAK.

In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors.

For additional comments, please visit our web site and read our Publication ethics statement (https://eujap.uniri.hr/publication-ethics/). To get a sense of the review process and how the referee report ought to look like, the prospective Authors are directed to visit the *For Reviewers* page on our web site (https://eujap.uniri.hr/instructions-for-reviewers/).

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style
- contain footnotes and no endnotes
- contain references in accordance with the author-date Chicago style, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book
T: (Nozick 1981, 203)
R: Nozick, R. 1981. *Philosophical Explanations.* Cambridge: Harvard University Press.

Book with multiple authors

T: (Hirstein, Sifferd, and Fagan 2018, 100)

R: Hirstein, William, Katrina Sifferd, and Tyler Fagan. 2018. *Responsible Brains: Neuroscience, Law, and Human Culpability*. Cambridge, Massachusetts: The MIT Press.

Chapter or other part of a book
T: (Fumerton 2006, 77-9)
R: Fumerton, Richard. 2006. 'The Epistemic Role of Testimony: Internalist and Externalist Perspectives'. In *The Epistemology of Testimony*, edited by Jennifer Lackey and Ernest Sosa, 77–91. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199276011.003.0004.

Edited collections
T: (Lackey and Sosa 2006)
R: Lackey, Jennifer, and Ernest Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.

Article in a print journal
T: (Broome 1999, 414-9)
R: Broome, J. 1999. "Normative requirements." *Ratio* 12: 398-419.

Electronic books or journals
T: (Skorupski 2010)
R: Skorupski, John. 2010. "Sentimentalism: Its Scope and Limits." Ethical Theory and Moral Practice 13 (2): 125–36. https://doi.org/10.1007/s10677-009-9210-6.

Article with multiple authors in a journal
T: (Churchland and Sejnowski 1990)
R: Churchland, Patricia S., and Terrence J. Sejnowski. 1990. "Neural Representation and Neural Computation." *Philosophical Perspectives 4*. https://doi.org/10.2307/2214198

T: (Dardashti, Thébault, and Eric Winsberg 2017)
R: "Dardashti, Radin, Karim P. Y. Thébault, and Eric Winsberg. 2017. Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity." *The British Journal for the Philosophy of Science* 68 (1): 55–89. https://doi.org/10.1093/bjps/axv010

Website content
T: (Brandon 2008)
R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013. http://plato.stanford.edu/archives/fall2010/entries/natural-selection

Forthcoming
For all types of publications followed should be the above guideline style with exception of placing 'forthcoming' instead of date of publication. For example, in case of a book:
T: (Recanati forthcoming)
R: Recanati, F. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material
T: (Gödel 1951)
R: Gödel, K. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications*. Unpublished manuscript, last modified August 3, 1951.

Final proofreading

Authors are responsible for correcting proofs.

Copyrights

The journal allows the author(s) to hold the copyright without restrictions. In the reprints, the original publication of the text in EuJAP must be acknowledged by mentioning the name of the journal, the year of the publication, the volume and the issue numbers and the article pages.

Archiving rights

The papers published in EuJAP can be deposited and self-archived in the institutional and thematic repositories providing the link to the journal's web pages and HRČAK.

Subscriptions

A subscription comprises two issues. All prices include postage.

Annual subscription:

International:

individuals € 30

institutions € 50

Croatia:

individuals 100 kn

institutions 375 kn

Bank: Zagrebačka banka d.d. Zagreb

SWIFT: ZABAHR 2X
IBAN: HR9123600001101536455

Only for subscribers from Croatia,

please add: “poziv na broj”: 0015-03368491

European Journal of Analytic Philosophy is published twice per year.

The articles published in the European Journal of Analytic Philosophy are indexed and abstracted in SCOPUS, The Philosopher's Index, European Reference Index for the Humanities (ERIH PLUS), Directory of Open Access Journals (DOAJ), PhilPapers, and Portal of Scientific Journals of Croatia (HRČAK), ANVUR (Italy), Sherpa Romeo