The Harmful-Dysfunction Account of Disorder, Individual versus Social Values, and the Interpersonal Variability of Harm Challenge

Antoine C. Dussault

Collège Lionel-Groulx; Centre interuniversitaire de recherche sur la science et la technologie.

This is a preprint of an article published in *Medicine, Health Care and Philosophy*. The final authenticated version is available online at: https://doi.org/10.1007/s11019-021-10021-8

Abstract

This paper presents the *interpersonal variability of harm* challenge to Jerome Wakefield's harmful-dysfunction account (HDA) of disorder. This challenge stems from the seeming fact that what promotes well-being or is harmful to someone varies much more across individuals than what is intuitively healthy or disordered. This makes it at least prima facie difficult to see how judgments about health and disorder could, as harm-requiring accounts of disorder like the HDA maintain, be based on, or closely linked to, judgments about well-being and harm. This *interpersonal variability of harm* challenge is made salient by the difficulty faced by harm-requiring accounts of disorder in dealing satisfactorily with cases of intuitively disordered conditions that seem harmless because they do not deprive the individuals that they affect of anything that they value (e.g., desired infertility). I argue that this challenge is made more serious for the HDA by some clarifications Wakefield has recently made on harm. In recent publications, Wakefield dissociates himself from the sheer cultural-relativist view of harm attributed to him by some critics based on his linkage of harm to social values, and adopts a more qualified social-values-based view of harm that leaves room for criticism of the values endorsed by members of a cultural group at a given time. I show how Wakefield's qualified view makes it more difficult for the HDA to deal with the *interpersonal variability of harm* challenge, at least when applied to a Western cultural context where a high value is placed on autonomy and individual choice.

Keywords: Health; disorder; harmful-dysfunction account; Jerome Wakefield; harm; social values.

1. Introduction

Jerome Wakefield's *harmful dysfunction account* (HDA) defines *medical disorder* in relation to both the biological concept of dysfunction and the prudential concept of harm

(e.g., Wakefield 1992; Wakefield 2014). In so doing, it adopts a hybrid position within the classical divide in the philosophy of medicine between purportedly objective and science-based *naturalist* accounts of disorder, and openly value-laden and social-constructivist *normativist* accounts. Medical disorder, according to the HDA, is defined by two criteria: (1) it involves a dysfunction, and (2) it causes harm to its carrier.[1]

Wakefield supports the superiority of hybrid accounts like his own over strictly naturalist ones like Christopher Boorse's biostatistical theory (BST) (Boorse 1977; 1997; 2014), by pointing to an important seeming limitation of the latter accounts.[2] Those accounts, he claims, by considering dysfunction as sufficient for disorder, include among disorders dysfunction-involving conditions that are benign (i.e., cause no further trouble to their carriers), and that, for this reason, are (allegedly) considered healthy by lay people and medical professionals.[3] Among such conditions are the lack of one kidney after transplantation, the asymptomatic carriage of infectious agents, asymptomatic HIV infection, the carriage of neutral, risky or benign dysfunctions, *situs inversus totalis*, benign angiomas, etc. (Wakefield 2014, sec. 4; Wakefield and Conrad 2020, 357–359; Wakefield 2021a, 518–520). Since a common feature of these benign conditions is that they cause no harm to their carriers, Wakefield maintains that they show the superiority of hybrid accounts of disorder like the HDA over strictly naturalist ones like the BST. This *benign-dysfunction* challenge, indeed, does not definitively resolve the debate between hybrid and strictly naturalist accounts of disorder, but it surely adds weight in the balance in support of the former.

This paper aims to highlight a challenge raised by the HDA's harm criterion, which, I think, pulls the balance back in favor of strictly naturalist over hybrid accounts of disorder. This challenge arises from some clarifications Wakefield makes in recent publications of how he thinks the harm criterion in the HDA should be understood (Wakefield 2013; 2021b; Wakefield and Conrad 2019; 2020). A particularity of Wakefield's take on the harm criterion is his linkage of harm with social values. According to him, "a disorder exists when the failure of a person's internal mechanisms to perform their functions as designed by nature impinges harmfully on the person's well-being *as defined by social values and meanings*" (Wakefield 1992, 373, emphasis added). This *social-values-based* view of harm contrasts with possible alternatives that would

[1] I follow Wakefield here in using the term "disorder" to refer to the technical medical concept that he and other philosophers of medicine purport to analyze. The *disorder* concept is intended to include any condition that medicine regards as a departure from health, that is, not only conditions usually called "diseases," but also other types of departures, such as injuries, poisonings, growth disorders, etc. (see Wakefield 2014, 653). Other authors (e.g., Christopher Boorse) refer to this concept with the term "pathology."

[2] A noteworthy aspect of the HDA is Wakefield's characterization of the notions of function and dysfunction involved in the HDA's dysfunction criterion along the lines of the *selected-effects* theory of function advocated by many philosophers of biology (e.g., Millikan 1989; Neander 1991; Godfrey-Smith 1994). I will not be concerned with this aspect of the HDA here.

[3] In the present discussion, I will follow Wakefield and adopt a broadly understood method of conceptual analysis that focuses on the concepts of health and disorder that implicitly underlie medical professionals' and lay people's thinking about health and disease. Although this method is controversial (see e.g., Lemoine 2013; Schwartz 2014), I will adopt it here to locate my discussion in the same methodological space as that in which Wakefield locates his defense of the HDA.

define harm in relation to individuals' personal stance on their conditions, and/or in relation to purportedly objective or universal values (i.e., values whose validity transcends particular cultures). One rationale for the adoption of a social-values-based view of harm, as opposed to a view based in individual patients' values, is that, as Wakefield and Jordan Conrad note, "medicine is a socially sanctioned profession that carries with it a corresponding obligation to alleviate harm as judged by society" (Wakefield and Conrad 2019, 1; see also Wakefield 2021b, 557). An issue seemingly raised by Wakefield's social-values-based view of harm, however, as highlighted in recent critiques of the HDA, is that it seems to boil down to an unqualified *cultural-relativist* view of harm, which would problematically force the HDA to exclude conditions that clearly seem disordered and harmful from the class of disorders, simply because they are socially valued in some cultural contexts (e.g., anorexia in "pro-ana groups," i.e., groups who value anorexia and consider it healthy, see Feit 2017, 370–371; Cooper 2021, 538–539).[4] In recent publications where he discusses these criticisms, Wakefield elaborates upon how he thinks the linkage between harm and social values should be construed, and, importantly, shies away from the sheer cultural-relativist view of harm the critics attribute to him (Wakefield 2013; 2021b; Wakefield and Conrad 2019; 2020). In these publications, Wakefield adopts a view of harm as *qualifiedly*, rather than *unqualifiedly*, based in social values, which, as we shall see, leaves room for criticism and assessment of the values endorsed by members of a cultural group at a given time.

I will argue that Wakefield's more qualified social-values-based view of harm makes his HDA more vulnerable to another challenge, one that is commonly faced by *harm-requiring* accounts of disorder (i.e., accounts that define disorder fully or, like the HDA, partly in relation to harm). The challenge is that what intuitively promotes well-being or is harmful varies much more across individuals than what is intuitively healthy or disordered, such that it is, at least prima facie, difficult to see how judgments about the latter could be based on, or closely linked to, judgments about the former (see Hausman 2016, sec. 5). For instance, whether infertility is harmful seems to depend upon whether the individual it affects does or does not desire to conceive children, and therefore seems bound to vary across individuals, whereas whether infertility is a disorder does *not* seem to vary in this manner. Proponents of harm-requiring accounts of disorder thus need to explain how interpersonally consistent diagnostic judgments, that is, interpersonally consistent judgments about the healthy/disordered character of a condition, can be based on interpersonally variable assessments of the harmless/harmful character of that condition. I will call this challenge the *interpersonal variability of harm* challenge to harm-requiring accounts of disorder.

An *unqualified* cultural relativist view of harm of the kind Wakefield shies away from, it seems, would have a straightforward way to deal with this challenge: whether a dysfunction-caused inability (e.g., infertility) is harmful or not depends simply on whether one's culture disvalues it, not on whether one personally disvalues it. I will argue

[4] A similar example sometimes brought out in the philosophy of medicine literature is that of bound feet in pre-20th-century China, which was socially valued, but was presumably nevertheless a disordered condition (see Schramme 2002, 62; Kingma 2017, 11).

below that the more *qualified* social-values-based view of harm recently adopted by Wakefield makes it more difficult for the HDA to meet this challenge. On this view of harm, as I will highlight, a condition's harmfulness/harmlessness has to do, not with how members of one's culture evaluate this condition, but rather with what assessments of the pro tanto harmful/harmless character of that condition are justifiable based on one's culture's value system upon critical weighing of its various components. I will argue that such a view of harm generates problems for the HDA, at least as applied to a Western cultural context.

My discussion will be organized as follows. In section 2, I will introduce the *interpersonal variability of harm* challenge and illustrate it with two examples: desired infertility, and meat allergy affecting a committed vegan person. In section 3, I will summarize the main aspects of the *qualified* social-values-based view of harm recently adopted by Wakefield, and explain how he thinks such a view accounts for the process that led to the depathologization of homosexuality in the 1970s. In sections 4 and 5, I will discuss Wakefield's qualified social-values-based view of harm in relation to two interpretations of Western culture's value system, and show what they imply regarding conditions like desired infertility and meat allergy affecting a committed vegan. The first one—the *generalizing* interpretation—focuses on what *general* value judgments are justified based on the culture's value system, and the second one—the *autonomy-emphasizing* interpretation—emphasizes the importance of self-determination and individual choice in Western culture. In section 6, I will outline some implications of the *interpersonal variability of harm* challenge for the general debate between hybrid accounts of disorder like the HDA, and strictly naturalist accounts like the BST.

2. The interpersonal variability of harm challenge

A common type of counterexample raised against harm-requiring accounts of disorder is that of intuitively disordered conditions that seem harmless because they do not deprive the individuals that they affect of anything that they value. A well-known case of this type is (dysfunction-caused) infertility affecting one who does not want to have children (Boorse 1975, 53; Schramme 2002, 62; Hausman 2016, 28–29). Infertility seems to be a disorder even in this case. However, it seems difficult to see how it might be (even pro tanto) harmful in this case. It rather seems beneficial in that it frees the affected person from concerns over contraception and unwanted pregnancies.

Another relevant case is that of a committedly vegan person who has acquired an allergy to meat. Allergy to meat is known to be caused by the "lone star tick," a tick that carries a sugar called "alpha gal," which humans do not naturally have in their bodies, but which is present in red meat, pork and some milk products (see Sullivan 2014). When it bites, the tick injects alpha gal into a person's bloodstream and the person's immune system reacts by creating antibodies. The next time the person eats meat, the antibodies are activated and cause an allergic reaction (which can involve hives, breathing problems, blood pressure drops, and even anaphylactic shock). If we suppose, with Wakefield (1999, 466), that allergies involve dysfunctions and are disorders, then such allergies will constitute an intuitively disordered dysfunction-involving condition that is seemingly harmless when affecting a committedly vegan person. A committed vegan who is bitten and injected with alpha gal by a lone star tick, and, as a result, develops antibodies that

make her allergic to meat will, it seems, suffer no *pro tanto* harm from her condition. As a committed vegan, she would not eat meat under any circumstances, and the allergy would therefore make no difference to her life.[5] The committed vegan person might even welcome her allergic condition as a safeguard of her moral commitment not to eat meat in eventual episodes of *akrasia*.[6]

Counterexamples of this type illustrate what appears to be an important difference between assessments of well-being and harm, on the one hand, and assessments of health and disorder, on the other. Daniel Hausman (2016, sec. 5) highlights this difference in a review on the relationship between health and well-being. The difference is that what intuitively promotes well-being or is harmful varies much more across individuals than what is intuitively healthy or disordered. Hausman spells out this difference with respect to well-being and health:

> [A]mong people of the same sex and roughly the same age there is comparatively little variation in what counts as good health, while utterly different lives may be good lives and good, in part, because of their differences rather than despite them. To exaggerate the point, one might say that there is one way to be healthy, while there are many ways to have a good life. The good life for some people consists in taking risks, whereas others thrive in quiet comfort. Some people flourish by pursuing their ambitions, but others focus on friends and family. (Hausman 2016, 32)

As Hausman notes, this difference is partly explained by another one, namely, the (seeming) fact that what promotes one's well-being depends heavily upon one's personal goals and values, while personal goals and values are much less relevant to assessments of people's health. My above examples of desired infertility and meat allergy affecting a committed vegan clearly illustrate this second difference: whether infertility and meat allergy are harmful seems to depend on the affected person's goals and values, but whether they are disorders does *not* seem to similarly depend on her goals and values. The difference in the respective degrees of interpersonal variability that characterize assessments of well-being and harm, on the one hand, and assessments of health and disorder, on the other, has an important corollary (also highlighted by Hausman). A third difference between well-being and harm, on the one hand, and health and disorder, on the other, is that interpersonal well-being comparisons pose a significant challenge, while interpersonal comparisons of health seem as straightforward as intrapersonal comparisons (Hausman 2016, 32–33). It is a commonplace observation that closely similar circumstances may make one person happy and another one miserable. This partly explains why a non-controversial method for interpersonal well-being comparisons is difficult to establish (for discussions, see Hausman 1995; Rossi 2011). Interpersonal comparisons of health conditions are much less problematic. When we compare two health conditions, it does not seem to matter whether the conditions compared are borne

[5] Here, I assume the validity of the position recently adopted by the Academy of Nutrition and Dietetics: "Appropriately planned vegetarian, including vegan, diets are healthful, nutritionally adequate, and may provide health benefits for the prevention and treatment of certain diseases. These diets are appropriate for all stages of the life cycle, including pregnancy, lactation, infancy, childhood, adolescence, older adulthood, and for athletes" (see Melina et al. 2016).

[6] Along similar lines, harm-requiring accounts of disorder would seem to imply that a committed vegan who is allergic to honey, or to seafoods like shrimp, lobster and crab, would not be disordered.

by two distinct individuals (of the same sex and age group), or by a single individual at different times.

These differences between assessments of well-being and harm, on the one hand, and assessments of health and disorder, on the other, with regard to their respective degrees of interpersonal variability, raises a seeming challenge for accounts that (fully or partly) define the latter in terms of the former. Prima facie, it seems difficult to see how interpersonally variable assessments of well-being and harm could possibly ground interpersonally consistent assessments of health and disorder. I call this challenge seemingly faced by harm-requiring accounts of disorder the *interpersonal variability of harm* challenge.[7] This challenge creates the need for proponents of such accounts either to explain how interpersonally variable assessments of well-being and harm can yield interpersonally consistent or standardized assessments of health and disorder, or to explain away the challenge, possibly by disputing the idea that assessments of health and disorder must be consistent across individuals.[8]

What Wakefield says in response to proposals to define disorder in relation to the *personal values* of the individuals whose health condition is assessed (instead of to *social values*) indicates that he is looking for an account that yields interpersonally standardized assessments of health and disorder. In response to such proposals by Andreas DeBlock and Jonathan Sholl (2021, 497) and Rachel Cooper (2021, 545–547), he states:

> [T]he "individual harm" approach would make a hash of diagnosis for a variety of reasons. For starters, people's values change over time and sometimes within a short span. … A physician's job in diagnosis is not to psychoanalyze the patient and decide what the patient *really* wants or to discern whether the patient might change their mind the next day or ten years hence and more generally what the patient might want in the future. (Wakefield 2021a, 513, emphasis in the original)

As I mentioned in the introduction, one rationale for Wakefield's use of a social-values-based understanding of harm is that "medicine is a socially sanctioned profession that

[7] I should note that Hausman's point in highlighting the above differences between assessments of well-being and assessments of health is slightly different from mine. Hausman's point is to show that health cannot be considered as a kind of well-being, while my point is to argue that health (and disorder) cannot be defined (fully or partly) in terms of well-being (and harm). I think the differences he highlights nevertheless support my point. I should also note that Hausman (2016, 33) highlights a fourth difference between well-being and health: assessments of well-being, he claims, concern a person's whole life, whereas assessments of health concern a person's condition during some period. I leave this fourth difference aside here.

[8] The view that assessments of disorder should be consistent or standardized across individuals is controversial. Some maintain that disorder judgments should align with individuals' personal assessments of their condition (e.g., Cooper 2002; 2021; De Block and Sholl 2021). Rachel Cooper (2002, 274) brings up the case, very similar to my above infertility and meat allergy cases, of an artist who becomes colorblind after a head injury, and who comes to consider his condition as preferable to color vision because it makes him more sensitive to textures and patterns. For Cooper, cases like this one illustrate that, contrary to what I presuppose (following Hausman), "one and the same condition can be pathological for one person but not for another." As I will soon show, unlike Cooper, Wakefield is committed to providing an account that does not allow assessments of disorder to vary across individuals, and so his HDA faces the *interpersonal variability of harm* challenge.

carries with it a corresponding obligation to alleviate harm as judged by society" (Wakefield and Conrad 2019, 1; see also Wakefield 2021b, 557). Wakefield's last remark indicates another motivation for the social-value-based view: in Wakefield's view, diagnostic judgment should, to some extent, be standardized. At first glance, the social-values-based definition of harm seems suited to providing such standardization. It sets some distance between what an individual personally values, and what counts as harmful to her. In so doing, it implies that whether a dysfunction-involving condition carried by someone is a disorder depends on whether her social group's value system entails that it is, rather than on whether she subjectively experiences it as harmful. Hence, at first glance, the social-values-based view seems able to yield assessments of the healthy/disordered status of a condition that are standardized across individuals from this social group.

For instance, Wakefield claims that, in a Western cultural context, infertility remains a disorder even when it affects someone who does not want to have children, because Western culture values the ability to reproduce (Wakefield and Conrad 2019, 1; Wakefield 2021b, 557). The diagnostic judgment that infertility is a disorder is thus standardized across individuals, and insulated from the vagaries of individuals' personal stances regarding reproduction. Along similar lines, one could argue that the ability to enjoy the taste of meat is socially valued in Western culture, such that the inability to do so ensuing from a meat allergy is a disorder irrespective of individuals' personal stances regarding meat consumption. Hence, a motivation for the adoption of a social-values-based, as opposed to an individual-values-based, view of harm is that the former seems to make the HDA better able to deal with the *interpersonal variability of harm* challenge.

In the following sections, however, I will argue that it is at best unclear whether the social-values-based view of harm as Wakefield construes it *really* enables his HDA to escape the *interpersonal variability of harm* challenge. Although it would arguably do so on an *unqualified* cultural relativist view of harm of the kind attributed to Wakefield by his critics, I will maintain that it much less clearly does so on the more *qualified* social-values-based view of harm Wakefield adopts in response to these critics (Wakefield 2013; Wakefield and Conrad 2019; Wakefield 2021b).

3. Wakefield's qualified social-values-based view of harm

Wakefield's early statements linking harm to social values were rather general. As he states, for instance, in the paper where he first introduced the HDA, "disorder lies on the boundary between the given natural world and the *constructed social world*; a disorder exists when the failure of a person's internal mechanisms to perform their functions as designed by nature impinges harmfully on the person's well-being *as defined by social values and meanings*" (Wakefield 1992, 373, emphasis added; see also Wakefield 1992, 374, 384; Wakefield 2007, 149–150). Arguably, such statements are up for two interpretations:

(1) A view of harm as *unqualifiedly* defined by social values: A sheer cultural-relativist view of harm according to which a condition counts as harmful just if it, as a matter of fact, is considered so in a given sociocultural group; and

(2) A view of harm as *qualifiedly* defined by social values: A qualified cultural-relativist view of harm that leaves some room for the rational examination and criticism of the values encompassed in a culture's value system.

Although his critics have (as mentioned above) tended to attribute to him the former view, Wakefield formally rejects it in some recent publications (Wakefield 2013; Wakefield and Conrad 2019; Wakefield 2021b). He states: "the social judgment that a condition is harmful may be based on misguided social values" (Wakefield 2013, 1), and "when I claimed that social values provide an essential filter for judgments of medical harm, I did not mean to assert absurdities such as that 'whatever is disvalued by a society should be rejected' or that 'any condition that a society values is valuable'" (Wakefield 2021b, 555).[9] Wakefield here adopts a *qualified* social-values-based view of harm along the lines of the second interpretation. In recent publications, he and Conrad make important clarifications as to how they conceive harm.

A first clarification they make is that the HDA's harm criterion is concerned with *pro tanto* rather than *all things considered* harms (Wakefield and Conrad 2019, 1–2; 2020, sec. III). That is, the harms that matter when assessing the healthy/disordered character of a condition include those that might be offset by some indirect benefits, in such a way that the person who suffers them might experience her situation as overall beneficial. For instance, a broken arm is a disorder because the loss of arm mobility is considered *pro tanto* harmful in our culture, and this is so irrespective of the potential benefits that one might gain from this condition (e.g., time off work, insurance payouts). Likewise, cowpox remains a disorder in the course of a smallpox epidemic even if it provides immunity to smallpox, because cowpox remains *pro tanto* harmful in this case.

A second, and very important, clarification they make concerns the possibility of criticizing the values endorsed by a cultural group at a given time. Wakefield and Conrad explain this possibility as follows:

> [S]ocial values or standards … are part of a cultural value system that has a complex multilayered structure and that is open to critical scrutiny and revision in the course of a dialectic about which of a culture's many often-conflicting value commitments are most basic, how to adjudicate between competing values, whether some seeming values are really just rationalisations of unjust power or blind prejudice, and how changing circumstances should alter these judgements. (Wakefield and Conrad 2019, 1)

And slightly differently:

> Social values … are not initial superficial subjective reactions but value claims that have been subjected to a dialectic that goes deeper than immediate reactions or consensus to explore which of a culture's many often-conflicting value commitments are its most basic values, which serve long-

[9] Wakefield develops his more considered view of harm partly in response to criticism by Russell Powell and Eric Scarffe (2019), and Rachel Cooper (2021), pressing him to adopt a more objectivist view of harm in replacement for his social-values-based view. Wakefield remains skeptical about the objectivist view and motivates his preference for the social-values-based view (Wakefield and Conrad 2019, 2–3; Wakefield 2021b, 555–559). I will not discuss the implications of integrating an objectivist view of harm for the HDA, but I think it would raise essentially the same issues that I will raise below for Wakefield's qualified social-values-based view.

> run interests of justice, which might be reactions that rationalize power relations, and so on. (Wakefield 2021b, 555)

As Wakefield emphasizes, "no simple reduction of a social value system to a poll of the people in a society can explain such dynamic phenomena" (Wakefield 2021b, 556). His qualified social-value-based view of harm, he claims, thus provides a way to accommodate the idea that actual people and even whole cultures at a given time "can err about what is harmful," with no need to allude to some "realm of culture-transcendent moral values" (Wakefield 2021b, 555–556). One can "seek redress [for the potential moral errors of a culture] in the potential for moral change that exists within the resources and complexities of any actual human culture's value system." (Wakefield 2021b, 556)

Wakefield illustrates how a critical dialectic of the kind described in the above quotes was at play in the process that led to the depathologization of homosexuality in the 1970s (Wakefield 2013, 2; Wakefield 2014, 675–676; Wakefield and Conrad 2019, 2; Wakefield 2021b, 556, 564–565). Homosexuality was depathologized, Wakefield maintains, because it became apparent to experts that there were no grounds for considering homosexuality as *in itself* harmful, and that the harms experienced by homosexual people primarily resulted from the social ostracism they were victims of.[10] The experts' judgment was based on considerations pertaining to the fact that a homosexual orientation does not *in itself* affect one's access to aspects of human life considered valuable in Western culture. Those considerations, Wakefield (2021b, 564–565) notes, ranged from "lack of distress or role impairment to the lessening importance of childbearing in an overpopulated world and the primary importance of the ability to have loving adult relationships." These were "culturally anchored considerations," but "they were edgy and pushed the culture beyond immediate reactions to confront foundational value issues in a value dialectic." (Wakefield 2021b, 565) This value dialectic, he further explains, led already existent cultural values, such as equality and acceptance of others, to be extended to new features and individuals. The relevant shift in values here, Wakefield emphasizes, did not amount simply to a change in public opinions or attitudes towards homosexuality (Wakefield and Conrad 2019, 2). It was, if I understand correctly, a shift in what was seen as justifiable based on the culture's value system upon critical weighing of its various components. The medical experts' judgment was based on their culture's value system, but they looked beyond current socially dominant values, and focused on determining what value judgments were consistent with the core components of the culture's value system.

Our consideration of Wakefield's take on the depathologization of homosexuality in the previous paragraph anticipated on a third clarification made by Wakefield about

[10] Indeed, its dysfunction criterion might provide another ground on which the HDA could potentially explain the depathologization of homosexuality. Wakefield could set aside the question of whether their sexual orientation is harmful to homosexual people, and maintain that the homosexual orientation is likely not caused by a dysfunction in the selected-effects sense (for support for such a claim, see Lewens 2015, 187). Wakefield, however, commits himself to defending the view that homosexuality would not be a disorder *even if* it were caused by a (selected-effects) dysfunction. Hence, for the sake of the present discussion, I will follow Wakefield in assuming that what is a stake with regard to the HDA's implications for homosexuality is whether homosexuality is harmful.

harm. As we have just seen, the harms that he considers relevant to the assessment of the healthy/disordered character of a condition are those that justify considering the condition as *in itself* harmful, as opposed to more *indirectly* harmful (as in the case of the harms resulting from the social ostracism suffered by homosexual people). Wakefield and Conrad (2019, 1–2) explain this idea as follows: "not all harm caused by a dysfunction via any causal route is relevant; only harm that is the direct or the intrinsic result of the dysfunction qualifies for the HDA's 'harm' component." This excludes harms "resulting from society's reaction to a dysfunction." (see also Wakefield 2021a, 513–514)

One must be careful, however, when applying this *directness* criterion, because Wakefield in fact *does* want to include some indirect and socially-mediated harms among those that qualify a condition as a disorder. This is most apparent in his discussion of dyslexia. Wakefield argues that dyslexia (assuming it is caused by a dysfunction) is a disorder "not simply because 'culture A values being able to read' but because reading is crucial to accessing the educational, occupational, recreational, and informational resources of (our) culture A." (Wakefield 2021a, 513–514)[11] With regard to dyslexia, it is the *indirect* (and socially-mediated) effects that Wakefield considers the most significant ones:

> In a society as dependent on reading as ours, with multiple opportunities and resources from occupational to recreational activities dependent on the ability, someone incapable of learning to read and thus incapable of accessing such resources is considered to be harmed pro tanto even if she claims not to value reading. She is no more unharmed just because of her disclaimer than someone without legs who says they don't care about walking. (Wakefield 2021a, 512)

Wakefield also recognizes harms that result from culturally-variant "social roles and role expectations" as relevant to the assessment of a condition's healthy/disordered character (see Wakefield 2021b, 559).

So in both the homosexuality and dyslexia cases (I leave aside the *suitability for social roles* case, about which Wakefield says little), there are *direct* socially disvalued effects—i.e., the inability to be attracted to people of the opposite sex, and the inability to enjoy the pleasure of reading; and *indirect* socially-mediated harms—i.e., harms due to heteronormative biases in the society, and deprivations of opportunities that our society's way of life makes valuable and which are accessible only (or more easily) to people who can read (for a similar point, see De Block and Sholl 2021, 497). Wakefield hence does not consider *all indirect harms* as irrelevant to the assessment of a condition's healthy/disordered character. Wakefield's nuanced view, if I understand correctly, is that the relevant difference between relevant and irrelevant socially-mediated harms has to do with whether the harms are caused by negative social evaluations *themselves*, or whether they are caused by the more general social context to which those social evaluations respond. In the case of homosexuality, the primary harms are caused by negative social judgments towards homosexuality themselves. In contrast, in the case of dyslexia, although some harms may be caused by negative social judgments towards the inability

[11] Wakefield considers that the dysfunction that causes dyslexia is not a disorder in preliterate, or eventually in future postliterate societies (Wakefield 2005, 89; Wakefield and Conrad 2019, 2; Wakefield 2021a, 517; Wakefield 2021b, 558).

to read, other harms, and arguably the most important ones, result from the more general social context that makes certain socially valuable goods accessible only (or more easily) to people who can read. Dyslexic people's limited access to these socially valuable goods does not result primarily from negative social judgments about dyslexia, but is rather an effect that is inherent to the inability to read when one has this inability in the social context characteristic of modern Western societies.[12]

Hence, on Wakefield's view of harm as *qualifiedly* defined by social values, the harms that do not count when determining whether a condition is healthy or disordered are not *all* the indirect and socially-mediated ones, but, more restrictively, those that are *caused* by certain negative social evaluations (presumably, those evaluations that are unjustified based on the culture's value system upon critical weighing of its various components).

As I mentioned in the introduction, Wakefield adopts the qualified social-values-based view of harm partly in response to criticisms targeting the HDA's harm criterion and focusing on cases of intuitively disordered conditions that are socially valued in some cultural contexts (such as anorexia in "pro-ana" groups, see Feit 2017, 370–371; Cooper 2021, 538–539). Wakefield's qualified view of harm affords the HDA resources for dealing with such purported counterexamples. As he argues with respect to the anorexia case,

> like just about all other human beings, pro-ana individuals presumably understand that death is a bad thing and should be avoided if possible. … They also understand that it is bad to lead an impoverished life … [T]hose who have a dysfunction causing their anorexic pursuits … suffer a variety of other harms easily recognized as direct *pro tanto* harms by the pro-ana members themselves, eventually possibly including, for example, such harms as pain, loss of mobility, fatigue, and, ironically, the inability to thus present one's desirable body to others in social interactions. (2021b, 561)

So there is an important motivation for Wakefield's adoption of a *qualified* (as opposed to *unqualified*) social-values-based view of harm.

In the following sections, however, I will argue that the *qualified* social-values-based view of harm has a cost: it makes it harder for the HDA to deal with the *interpersonal variability of harm challenge* to harm-requiring accounts of disorder highlighted in section 2, and with the associated type of counterexamples (e.g., desired infertility and meat allergy affecting a committed vegan). As I remarked, the *unqualified* social-value-based view of harm Wakefield shies away from in his recent publications had a straightforward way to deal with the *interpersonal variability of harm* challenge, because it implies that whether a dysfunction-caused inability is harmful to one depends on whether one's culture disvalues it, not on whether one personally disvalues it. As I will

[12] This reading of the socially-mediated harms associated with dyslexia could perhaps be challenged, since, arguably, the limited access to social goods experienced by dyslexic people is, to a not insignificant degree, an effect of social value judgments about what legitimizes reduced (or enhanced) access to resources and occupational or recreational activities. I will nevertheless grant, for the sake of the discussion, that there is a meaningful difference between the harms that result from ostracizing attitudes towards homosexuality and some of the harms associated with conditions like dyslexia.

now argue, it is at best unclear whether the qualified view of harm also enables the HDA to meet this challenge.

4. The generalizing interpretation of Western values

To determine whether the view of harm as *qualifiedly* defined by social values enables the HDA to meet the *interpersonal variability of harm* challenge, we need to determine whether it can yield interpersonally standardized and intuitively plausible assessments of the healthy/disordered character of conditions. The qualified social-values-based view of harm must imply that all (or at least most) intuitively disordered dysfunction-involving conditions are harmful, and that all (or at least most) intuitively healthy dysfunction-involving conditions (if they exist) are harmless. For simplicity, I will focus on whether the qualified social-values-based view counts intuitively disordered conditions as harmful in the context of Western culture's current value system. If my above reading of Wakefield's qualified view of harm is correct, the issue will therefore come down to what assessments of the pro tanto harmful/harmless character of conditions that illustrate the interpersonal variability of harm challenge, like desired infertility and meat allergy affecting a committed vegan, are justifiable based on Western culture's value system upon critical weighing of its various components. As the following discussion will indicate, it will be informative to consider the cases of desired infertility and meat allergy affecting a committed vegan in comparison with the homosexuality case as discussed by Wakefield (see above).

I will consider two possible interpretations of what is implied by Western culture's value system upon critical weighing of its various components: (1) The *generalizing* interpretation, which focuses on what *general* judgments (i.e., judgments that apply to all members of the culture) are justified based on the culture's value system (discussed in this section); and (2) The *autonomy-emphasizing* interpretation, which emphasizes the importance of self-determination and individual choice in Western culture (discussed in section 5).

The generalizing interpretation focuses on what *general* judgments (i.e., judgments that apply to all members of the culture) are justified based on the culture's value system. What is at issue is whether, *on a general basis*, conditions like infertility and meat allergy (or other intuitively disordered conditions with regard to which individual values vary) should be considered harmful to individuals upon critical weighing of the components of Western culture's value system.

At first glance, the generalizing interpretation of Western culture's value system seems to yield plausible results with regard to infertility and meat allergy. The essential difference between fertile and infertile people is, of course, that the former, but not the latter, can conceive children, and hence raise children that they have conceived. The essential difference between people who have and do not have meat allergies is, of course, that the latter, but not the former, can digest meat. Presumably, the judgment that the inability to conceive children and raise children that one has conceived is a pro tanto harm is justifiable based on Western culture's (current) value system upon critical weighing of its various components. And likewise, presumably, the judgment that the inability to digest meat is a pro tanto harm is justifiable based on Western culture's

(current) value system upon critical weighing of its various components. Western culture's value system arguably (currently) sees the ability to conceive children and to raise one's biological children as valuable, and (currently) sees the ability to enjoy meat-based meals as valuable. At least, I will grant these points at this stage, although, of course, many environmentalists and animal rights advocates would challenge them. Presumably, then, infertile people and meat-allergic people are deprived of abilities whose lack is rightfully considered as pro tanto harmful in (current) Western culture. Hence, even when one does not desire to have children or to eat meat, one nevertheless counts as better off with than without abilities to do so. The HDA thus implies that people who lack these abilities are disordered regardless of whether they value them.

At first glance the generalizing interpretation of Western culture adequately distinguishes these two cases from that of homosexuality. An obvious difference between homosexual and heterosexual people is, of course, that the latter, but not the former, are able to feel sexually and/or romantically attracted to people of the opposite sex. This inability of homosexual people has historically been considered as a ground for considering their condition as pro tanto harmful to them. The ability to be attracted to people of the opposite sex was socially valued, and, therefore, the inability to do so was considered a harm. This judgment, however, is unjustifiable based on Western culture's value system upon critical weighing of its various components. It involves a heteronormative view of sexual and romantic life that Western culture's value system no longer vindicates (this, as we have seen, is what Wakefield considers to have led to the depathologization of homosexuality). What is justifiable, perhaps, is the judgment that being unable to feel sexually and/or romantically attracted to *some* people is pro tanto harmful, but, clearly, not the judgment that the people one is attracted to should mandatorily belong to the opposite sex. Hence, at first glance, the qualified social-values-based view of harm, combined with the generalizing interpretation of Western values, implies that their sexual orientation does not deprive homosexual people of any ability whose lack is rightfully considered as pro tanto harmful in Western culture.

However, this reading of the generalizing interpretation of Western culture's implications with regard to homosexuality turns out to be too superficial if we bear in mind what I remarked above concerning Wakefield's nuanced use of the *directness* criterion for harm. As we have seen above in Wakefield's treatment of dyslexia, the *directness* criterion does not make *all* indirect harms irrelevant to assessments of the healthy/disordered character of a condition. Indirect harms may matter when they are caused not by negative social evaluations *themselves*, but by the more general context to which those social evaluations respond. But if this is the case, then the above reading of the generalizing interpretation's implications regarding homosexuality is too superficial in that it considers only the (non-harmful) inabilities that directly result from homosexuality, and overlooks some losses of opportunity that more *indirectly* ensue from it. Considering the indirect losses of opportunity in fact brings the homosexuality and the infertility cases closer to each other than is implied by the above treatment. Independently of any negative social judgments about homosexuality, homosexual people are deprived of opportunities to conceive children in a sexual act with their loved one, and to raise children of whom they and their loved one are the two biological parents. These opportunities seem to be considered valuable in Western culture's (current) value system.

These opportunities in fact seem to partly explain the value ascribed to fertility in Western culture, and the ensuing consideration of infertility as harmful. *Hetero*sexual couples who cannot conceive children together, and hence cannot raise children of whom they and their loved one are the two biological parents, are generally considered in a less enviable situation than couples who can. So on an interpretation of Western culture's value system that looks for *general* value judgments (i.e., judgments that apply to all members of the culture), consistency would require also considering *homo*sexual couples as in a less enviable situation than couples who can conceive children together and raise children of whom they and their loved one are the two biological parents. This implies that homosexual people's lack of these opportunities is harmful to them, as much as infertility (purportedly) is to infertile people. The qualified social-values-based view of harm, combined with the generalizing interpretation of Western values, therefore problematically commits the HDA to the idea that homosexuality is a disorder.[13]

Hence, when indirect losses of opportunities, besides direct inabilities, are taken into account, the generalizing interpretation of Western culture's value system makes it difficult for the HDA to reconcile the idea that infertility is a disorder with the idea that homosexuality is not. Either the ability to conceive children with one's loved one is valuable, and then both infertility and homosexuality are harmful and therefore disorders, or it is not, and then both infertility and homosexuality are harmless and therefore not disorders. Wakefield himself, as seen in the above quote (see section 3), includes among the considerations that led to the depathologization of homosexuality, "the lessening importance of childbearing in an overpopulated world" (Wakefield 2021b, 564; see also Wakefield 2014, 676). This seems to commit him to the view that infertility is harmless, which would in turn commit the HDA to the implausible view that infertility is not a disorder. Interestingly, Wakefield (1992, 384) recognizes that the idea that the ability to have children is a benefit and that its deprivation is a disorder "has been disputed because of its implications for the classification of homosexuality." It is unclear, however, that disputing the assumed harmfulness of infertility comes together with disputing its seemingly disordered character. It rather seems that (dysfunction-caused) infertility would remain a disorder even in a socio-cultural context where it was deemed harmless.

Wakefield's qualified social-values-based view of harm, combined with the generalizing interpretation of Western culture's value system, admittedly yields interpersonally standardized diagnostic judgments with regard to infertility. However, it yields ones that are implausible. Wakefield could possibly avoid this implication by making stricter use of his *directness* criterion, and include only *strictly direct* harms among those that qualify a dysfunction-involving condition as a disorder. Then, infertility would count as a disorder under the HDA because it *directly* causes an inability to reproduce, whereas homosexuality would *not* count as a disorder under the HDA because it only *indirectly* causes a loss of opportunities. This move, however, would undermine

[13] Of course, this is the case only if we assume that homosexuality involves a (selected-effects) dysfunction, and whether it really does so is debatable (see footnote 10 above). As I said, for the sake of the present discussion, I follow Wakefield in assuming that what is a stake with regard to the HDA's implications for homosexuality is not whether homosexuality involves a dysfunction, but whether it is harmful.

the HDA's ability to consider dyslexia and other conditions that primarily cause socially-mediated harms (e.g., ones tied to unsuitability for social roles) as disorders.

The current evolution of Western values regarding meat eating may prompt a reconsideration of the idea that meat allergy is harmful similar to the one that, as Wakefield remarks, was prompted with regard to infertility and homosexuality by the evolution of Western values regarding childbearing. As equally nutritious and enjoyable vegan alternatives to animal-derived ingredients develop, and as the ecological damage and animal harms caused by livestock farming are recognized, Western culture may well end up no longer ascribing value to the ability and opportunity to enjoy meat-based meals, and hence no longer consider the inability to digest meat as harmful. I doubt, however, that such an evolution of Western values regarding meat-based food would have any bearing on the healthy/disordered status of allergy to meat. It rather seems that an allergy to meat, simply by virtue of being an allergy, would remain a disorder even in a socio-cultural context where it was deemed harmless.

Hence, the generalizing interpretation of Western culture's value system makes it difficult for the HDA to meet the *interpersonal variability of harm* challenge. On this interpretation, the HDA can meet this challenge only if the general value judgments that ensue from the culture's value system upon critical weighing of its components imply that all (or at least most) intuitively disordered conditions are harmful, and that all (or at least most) intuitively healthy conditions are harmless. We have seen, however, that it is at best uncertain whether this criterion can be met at the same time for infertility, homosexuality, dyslexia, and meat allergy.

5. The autonomy-emphasizing interpretation of Western values

Another possible interpretation of what Western culture's value system implies regarding harm emphasizes the importance of self-determination and individual choice within it, and takes as central to it the idea that what is pro tanto beneficial or harmful to someone depends chiefly on her own perspective on her life. Under this interpretation, Western culture's value system defines general goods that members of the culture can ascribe value to, but also allows that not *all* individuals ascribe value to *all* these goods and that the particular content given to these goods varies according to individuals' preferences. Although the culture's value system still makes some harm judgments illegitimate (e.g., presumably, the judgment that anorexia causes no pro tanto harm), it nevertheless allows that, with respect to many conditions, what is beneficial or harmful differs across individuals.

On the face of it, this interpretation of Western culture's value system seems more plausible than the previous one. It is rather strange to assert that the inability to conceive children and the inability to digest meat are or are not pro tanto harmful to people irrespective of whether they do or do not care about these inabilities, or about the lack of opportunities that ensue from them. The acceptance of variation among individuals with regard to what a *good life* consists in is a defining trait of Western culture, and it undergirds the importance that Western societies ascribe to individual freedom and autonomy. When the importance of self-determination and individual choice within Western culture is taken into account, the inference from "Western culture values X" to

"the inability and/or lack of opportunity to X is pro tanto harmful" seems hasty. Western culture also values individuals' personal takes on the abilities and opportunities that it values, and leaves it to a significant degree up to those individuals to decide whether these abilities and opportunities matter to them. Hence, from a Western cultural perspective, the lack of ability and/or opportunity to access certain culturally valued goods need not be pro tanto harmful to *all* members of the cultural group.

This interpretation of Western culture's value system has implications that differ from the generalizing one regarding the cases of infertility, meat allergy and homosexuality. It implies that infertility is pro tanto harmful to people who want to reproduce, but not to people who don't. It implies that homosexuality is, with regard to its implications for reproduction, indirectly pro tanto harmful only to homosexual people who value the opportunity to conceive children with their loved one and to raise children of whom they and their loved one are the two biological parents. And it implies that meat allergy is pro tanto harmful to meat eaters, but not to committed vegans. People who do not want to reproduce are legitimized to say that, for them, infertility is not, or would not be, a harm, because they envision other ways to nurture the next generation, some of which are equally valuable to them. Homosexual people are justified to consider that a non-conventional family is as adequate as a conventional one, such that their lack of opportunity to form a family through the same means as most heterosexual couples does not put them in a less enviable situation. And it is up to committed vegans who develop an allergy to meat to consider that, since a vegan diet may be as balanced and enjoyable as a meat-involving one, their allergy deprives them of nothing that they deem valuable.[14]

By embracing interpersonal variation from the outset, however, this interpretation of Western culture's value system makes it prima facie even more difficult for the HDA to meet the *interpersonal variability of harm* challenge. Without additional qualifications, it implies that, on the HDA, infertility is a disorder only when affecting people who *do* want to reproduce, and that meat allergy is a disorder only when affecting meat eaters. The autonomy-emphasizing interpretation of Western culture thus, without further qualifications, makes the HDA unable to yield interpersonally standardized diagnostic judgments. Against Wakefield's intentions (see section 2), the HDA would hence require physicians to base their diagnostic judgments partly on inquiries into what their patients *really* want and what they might want in the future. Moreover, the autonomy-emphasizing interpretation would still carry unsatisfactory implications regarding homosexuality: it would imply that homosexuality is a disorder in the case of homosexual people who value the opportunity to conceive children with their loved one and to raise children of whom they and their loved one are the two biological parents.

Wakefield, however, could possibly escape these implications by appealing to qualifications that he and Conrad introduce in response to recent criticism of the HDA's harm criterion (Muckler and Taylor 2020). Those qualifications set some conceptual distance between individuals' assessments of harms and what counts as harmful in the sense relevant for the HDA. Wakefield and Conrad (2020, 354–359) introduce two such

[14] I thank Anne-Marie Gagné-Julien for insights that reinforced the points made in this paragraph.

qualifications. They propose that a dysfunction-involving condition is a disorder according to the HDA if it is *typically* or *dispositionally* harmful (Wakefield and Conrad 2020, 354).[15]

The typicality qualification

Perhaps Wakefield could adopt the autonomy-emphasizing interpretation of Western culture and enable his HDA to yield interpersonally standardized diagnostic judgments by stipulating that, although pro tanto harmfulness varies across individuals, the HDA's harm criterion is concerned with *typical* pro tanto harmfulness rather than pro tanto harm affecting the particular individuals whose health condition is assessed. Wakefield and Conrad (2020, 354) adopt such a construal of the harm criterion, and apply it to the desired infertility case. On this construal of the harm criterion, infertility could count as a disorder even when affecting one who does not want to have children, because *most* infertile people consider their loss of ability as pro tanto harmful. Along similar lines, a proponent of the HDA could argue that meat allergy is a disorder even when affecting a committedly vegan person, because, presumably, *most* meat allergic people (who are not committed vegans) would consider their loss of ability as pro tanto harmful. Such a view linking disorder to *typical* harm, rather than to harm affecting the particular individual whose health condition is assessed, has been adopted by other proponents of harm-requiring accounts of disorder (e.g., Reznek 1987, 161–162).

At first glance, the *typicality* qualification appears to be a promising way for the HDA to deal with the *interpersonal variability of harm* challenge. By stipulation, it sets some distance between diagnostic judgments and particular individuals' assessments of the harmful/harmless character of their condition. It considers a dysfunction-involving condition as harmful in the sense relevant to the HDA, not when it is pro tanto harmful to the individual whose healthy/disordered condition is assessed, but instead when it is typically considered pro tanto harmful by members of her cultural group. The typicality qualification thus, even when combined with the autonomy-emphasizing interpretation of Western culture, enables the HDA to yield interpersonally standardized diagnostic judgments.

A possible worry, however, is that, as a stipulative way to obtain interpersonally standardized diagnostic judgments out of assessments of pro tanto harms that are acknowledged to be interpersonally variable, the *typicality* qualification amounts to a modification of the HDA, a modification that robs it of at least some of its intuitive appeal. The typicality qualification entails that a disorder is no longer a *harmful dysfunction*—a dysfunction that causes a harm—but, instead, a dysfunction that *would* be pro tanto harmful if affecting an individual whose personal assessment of her condition aligns with those of most members of her cultural group. On the *unqualified* social-values-based view of harm (which Wakefield shies away from in his recent publications),

[15] I am thankful to an anonymous referee for noticing the *typicality* and *dispositionality* qualifications. Wakefield and Conrad (2020) treat these qualifications as equivalent, but I will deal with them separately because I think they are distinct. Typicality may be understood in strictly statistical terms, whereas dispositionality cannot, because a disposition can have actualization conditions that rarely obtain and therefore be actualized infrequently.

and on the *qualified* view tied to the *generalizing interpretation* of Western culture's value system, disordered conditions were still ones that are pro tanto harmful *to the individuals carrying those conditions*. The harmfulness of disordered condition was *constituted* by the (justified) negative social judgments about it, not by the individuals' personal judgments. However, on the qualified view of harm tied to the *autonomy-emphasizing interpretation* of Western culture, individuals who find their condition perfectly fine, despite its being typically deemed pro tanto harmful by members of their cultural group, are no longer *individually* harmed by their condition. They have a condition that is pro tanto harmful to most members of their culture, but is harmless to them. Infertile or meat allergic people who are perfectly happy with their condition hence count as disordered, not by virtue of being pro tanto harmed by their condition, but by virtue of carrying a condition that most members of their cultural group deem pro tanto harmful. An account that considers one as disordered just when she carries a dysfunction that *would* be pro tanto harmful *if* carried by a representative member of her cultural group, however, is much less intuitively appealing than an account that considers one as disordered just when she carries a dysfunction that is pro tanto harmful *to her*. The shift from *harmfulness* to *typical harmfulness* thus has a high cost in intuitiveness.

Moreover, the typicality qualification still yields problematic results with regard to homosexuality. It makes the healthy/disordered status of homosexuality hinge on whether homosexual people *typically* consider their lack of opportunities to conceive children with their loved one, and to raise children of whom they and their loved one are the two biological parents, as losses. This seems too unsteady grounds for considering homosexuality as non-disordered. Perhaps it is the case that most homosexual people see no loss in the lack of these opportunities. However, a satisfactory account of disorder should, it seems, be able to say that, even if most of homosexual people *did* see the lack of these opportunities as losses, their sexual orientation would still not be a disorder.

The dispositionality qualification

Another possible way to enable the HDA to yield interpersonally standardized diagnostic judgments, even when combined with the autonomy-emphasizing interpretation of Western culture, is to introduce the qualification that harms that matter to the HDA might simply be ones that a dysfunction *disposes* one to suffer, and need not be ones that a dysfunction *actually* causes. For instance, a proponent of the HDA could argue that the conditions of the infertile person who does not want to reproduce, and of the meat-allergic committed vegan, *dispose* them to suffer harms, and that this is sufficient for the HDA to consider their conditions as disorders. Should the happily infertile person change her mind about reproduction, she would suffer harmful effects from her condition. Likewise, should the meat-allergic committed vegan accidentally eat meat products, she would suffer the harmful effects of her meat allergy. These are *dispositions to harm* that fertile and non-meat-allergic people are not disposed to. A proponent of the HDA could thus possibly maintain that, in *all* similar cases of intuitively disordered but seemingly harmless dysfunction-involving conditions, a *disposition* to harm is present.

The concept of disposition is a complicated one (see Choi and Fara 2018), but, roughly, an object *O* is considered to be disposed to produce the effect *E* under condition

C if and only if *O* would produce *E* if *C* obtained. Crucial to dispositions is thus their intimate linkage to *actualization conditions*, conditions under which the disposition should be actualized. When introducing the *dispositionality* qualification, however, Wakefield and Conrad (2020) do not specify what they take to be the relevant actualization conditions for dispositional harms. Presumably, they would want to adopt a rather restricted take on what those actualization conditions are. A too liberal take would have implications that diverge from Wakefield's stance on certain conditions, for instance, certain cases of asymptomatic infections and dyslexia. If all that were required for a dysfunction to dispose one to suffer harms was that there be *some possible conditions* under which this dysfunction would cause harms, then many conditions that Wakefield considers harmless and hence healthy would have to count as disorders. For instance, at least some asymptomatic infections by bacteria and viruses would have to count as disorders (Wakefield brings out this case, among other ones, to highlight the seeming insufficiency of dysfunction for disorder, and hence show the HDA's superiority over strictly naturalist accounts like Boorse's BST). Take Wakefield's example of infection by *Streptococcus pneumoniae*, a bacterium known as a major cause of pneumonia. As Wakefield states, this infection "does not always constitute a disorder because the vast majority of infections occur harmlessly in the nose and sinuses and the bacterium only becomes problematic under special circumstances, when it migrates to the lungs and becomes more virulent." (Wakefield 2021a, 520; see also Wakefield and Conrad 2020, 357–358) A liberal take on actualization conditions, however, would imply that infection by *Streptococcus pneumonia* always disposes one to harm, and would hence commit the HDA to the view that it is always a disorder. As Wakefield recognizes in the above passage, there *is* a condition under which an asymptomatic infection with *Streptococcus pneumonia* would cause harm: it *would* do so *if* the bacterium migrated to the infected person's lungs. Along similar lines, a liberal take on actualization conditions would imply that dyslexia in pre-literate societies, which Wakefield claims was not a disorder (Wakefield 2005, 89; Wakefield and Conrad 2019, 2; Wakefield 2021a, 517; Wakefield 2021b, 558), in fact *was* a disorder. A person with the dysfunction associated with dyslexia and living in a preliterate society *would* suffer harms from her condition *if* her society became a literate one. This, admittedly, is a rather fictitious scenario, but absent any specification of actualization conditions for dispositional harms, it is unclear what principled grounds Wakefield might have to exclude it.

So Wakefield and Conrad's implicit take on actualization conditions must be a more restrictive one. Some remarks they make when dealing with a case brought out by Muckler and Taylor (2020) may give us some indication of what their implicit take is. The case is that of a person with mild mononucleosis who, according to Muckler and Taylor, suffers no harm from her condition, because its only symptom is a tendency to fatigue and the person does not engage or even intend to engage in any physical exertion (Muckler and Taylor 2020, 337). This, in their view, illustrates the possibility of harmless disorders and so constitutes a counterexample to the HDA. Wakefield and Conrad respond that a problem with Muckler and Taylor's setting is that it artificially eliminates "expectable real-life contingencies … to which all people are exposed" and so is poorly informative of how the notion of harm works in more realistic situations (Wakefield and Conrad 2020, 354). They note that, even if the person does not intend to exert, "life inevitably requires exertion at times," for instance in contexts of "survival in running

from danger, protection of loved ones, success, romantic passion" (Wakefield and Conrad 2020, 354).[16] For these reasons, Wakefield and Conrad consider the person exemplified by Muckler and Taylor as *disposed* to suffer harms.

These remarks suggest the following take on actualization conditions for dispositions to harm: a dysfunction disposes one to suffer harm if "expectable real-life contingencies" make it sufficiently probable that this dysfunction will cause one some harms. This characterization of the actualization conditions for dispositional harms deals satisfactorily with the above case of nose-based or sinus-based asymptomatic *Streptococcus pneumonia* infection. If we assume, with Wakefield (2021a, 520), that "the vast majority of [*Streptococcus pneumonia*] infections occur … in the nose and sinuses," then it seems fair to consider that "expectable real-life contingencies" do *not* make it sufficiently probable that such infections will cause harms to their carriers. As Wakefield and Conrad (2020, 357) specify, the bacterium migrates to the lungs and becomes virulent only "under special circumstances, such as infection with influenza virus … or in an immunosuppressed host." The above characterization of actualization conditions also deals satisfactorily with the case of dyslexia in pre-literate societies. Expectable real-life contingencies do not make it probable that a person who carries the dysfunction associated with dyslexia and lives in a pre-literate society will come to find herself in a literate society. Non-literate societies do not suddenly become literate ones.

Now, the question is whether we should be confident that a similar disposition to harm will be present in *all* (or most) cases of intuitively disordered but seemingly harmless dysfunction-involving conditions like desired infertility and meat allergy affecting a committed vegan. I think it is at best dubitable whether it will. Under the more restrictive take on actualization conditions, it seems that happily infertile people will not necessarily be disposed to suffer harms from their condition. Indeed, people who do not want to have children sometimes change their minds. However, some don't, and society considers their lack of desire to have children as stable enough to recognize their eventual choice to acquire sterility through a medical intervention as reasonable. There is, it seems, no reason to reject the possibility of *naturally infertile* people with an equally stable lack of desire to have children. In their case, there seem to be no "expectable real-life contingencies" that make it sufficiently probable that they will be harmed by their condition. In other words, cases where infertility is stably desired seem, with regard to their dispositional harmful/harmless character, more similar to Wakefield's cases of nose-based or sinus-based *Streptococcus pneumonia* infection, than to Muckler and Taylor's case of mild mononucleosis. But happily infertile people, it seems, nevertheless have a disorder. Likewise for the meat allergy case. Suppose that a committed vegan takes stringent precautions in order to avoid consuming meat products, and does so, not because she is allergic to meat, but simply because of her commitment not to eat meat. Not only does she never buy meat products, but, also, she consents to live only with roommates who are as strongly committed vegans as her, and she eats only at restaurants or with friends who don't cook meat-based food. In this case, I submit that "expectable

[16] Here, I consider Wakefield and Conrad's remarks in abstraction from some details they give in order to locate their analyses within the context of the particular theories of well-being and harm discussed by Muckler and Taylor (e.g., the objective-list theory, hedonism, etc.).

real-life contingencies" do not make it probable that she will accidentally eat meat. If such a person happens to be bitten by a lone star tick and to develop a meat allergy, her condition will therefore not dispose her to suffer harms. But her condition, it seems, will nevertheless be a disorder.

Hence the dispositionality qualification, when combined with the autonomy-emphasizing interpretation of Western culture, does not seem to fully enable the HDA to satisfactorily deal with the *interpersonal variability of harm* challenge. Cases of intuitively disordered dysfunction-involving conditions that do not cause *dispositions to harm* seem possible. To be sure, this conclusion partly hinges on the interpretation I made of Wakefield and Conrad's (implicit) take on actualization conditions for dispositional harms, and perhaps a more refined characterization of those conditions could be elaborated. But the burden to provide such a more refined characterization is on their side.

Hence, it seems fair to conclude, at this stage, that the autonomy-emphasizing interpretation of Western culture's value system makes it difficult for the HDA to meet the *interpersonal variability of harm* challenge. Meeting this challenge seems to require proponents of the HDA to do one of two things. First, they may modify the HDA's classical definition of disorder as harmful dysfunctions, and adopt the less intuitive definition of disorder as dysfunctions that *would* be pro tanto harmful *if* affecting an individual whose personal assessment of her condition aligns with those of most members of her cultural group. Second, HDA proponents may link disorder to *dispositional* harm rather than exclusively to actual harms, and then make the HDA's validity hinge on their ability to provide a refined characterization of actualization conditions for dispositional harms. This refined characterization would have to ensure that dispositional harms are present in *all* (or most) cases of dysfunction-involving conditions that do not deprive the individuals that they affect of anything that they value. The autonomy-emphasizing interpretation of Western culture's value system thus leaves uncertain whether the qualified social-values-based view of harm recently adopted by Wakefield can yield interpersonally standardized diagnostic judgments, and so preserve the HDA's ability to meet the *interpersonal variability of harm* challenge.

6. Conclusion

Above, I argued that the *qualified* social-values-based view of harm that Wakefield adopts in response to recent criticisms makes his HDA more vulnerable to what I called the *interpersonal variability of harm* challenge, a challenge commonly faced by harm-requiring accounts of disorder. The challenge ensues from the seeming fact that what intuitively promotes well-being or is harmful varies much more across individuals than what is intuitively healthy or disordered, such that it is prima facie difficult to see how judgments about the latter could be based on judgments about the former. This, I maintained, creates the need for proponents of harm-requiring accounts of disorder, and for proponents of the HDA in particular, to explain how interpersonally consistent assessments of a condition's healthy/disordered character can be based on (interpersonally variable) assessments of their harmless/harmful character. I considered whether Wakefield's *qualified* social-values-based view of harm can meet this challenge in relation to two alternative interpretations of Western culture's value system: the

generalizing interpretation, and the *autonomy-emphasizing interpretation*. I maintained that, when combined with the generalizing interpretation, Wakefield's qualified social-values-based view of harm *does* yield interpersonally standardized diagnostic judgments, but ones that are implausible with regard to at least some important cases. And I maintained that, when combined with the autonomy-emphasizing interpretation, it is uncertain whether the qualified social-values-based view of harm can yield interpersonally standardized diagnostic judgments. This leaves it at best uncertain whether the HDA can meet the *interpersonal variability of harm* challenge.

Since, as we have seen, the *unqualified* social-values-based view of harm attributed to Wakefield by its critics also raises challenges, the above discussion suggests that the problem resides in the HDA's harm criterion itself. Admittedly, strictly naturalist accounts of disorder which include no references to harm also face challenges, and in particular, Boorse's biostatistical theory faces the above mentioned *benign-dysfunction* challenge raised against it by Wakefield (2014). Yet by highlighting the *interpersonal variability of harm* challenge, the above discussion nevertheless adds weight in the balance in favor of the latter type of accounts.

Acknowledgments

The author is thankful to Christopher Boorse, Anne-Marie Gagné-Julien, Christian Saborido, and two anonymous referees for very helpful comments that led to significant improvements of the paper. He also thanks Alice Everly for editing the manuscript. The work for this paper was supported by a research grant from the Fonds de recherche du Québec – Société et culture (FRQSC, 2018-CH-211053).

References

Boorse, Christopher. 1975. On the distinction between disease and illness. *Philosophy and Public Affairs* 5: 49–68.

Boorse, Christopher. 1977. Health as a Theoretical Concept. *Philosophy of Science* 44: 542–573.

Boorse, Christopher. 1997. A Rebuttal on Health. In *What Is Disease?*, ed. James M. Humber and Robert F. Almeder, 1–134. Totowa, N.J.: Humana Press.

Boorse, Christopher. 2014. A Second Rebuttal On Health. *Journal of Medicine and Philosophy* 39: 683–724.

Choi, Sungho, and Michael Fara. 2018. Dispositions. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2018. Metaphysics Research Lab, Stanford University.

Cooper, Rachel. 2002. Disease. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 33: 263–282.

Cooper, Rachel. 2021. On Harm. In *Defining mental disorder: Jerome Wakefield and his critics*, ed. Luc Faucher and Denis Forest, 537–551. Cambridge, Mass.: MIT Press.

De Block, Andreas, and Jonathan Sholl. 2021. Harmless Dysfunctions and the Problem of Normal Variation. In *Defining mental disorder: Jerome Wakefield and his critics*, ed. Luc Faucher and Denis Forest, 495–510. Cambridge, Mass.: MIT Press.

Feit, Neil. 2017. Harm and the concept of medical disorder. *Theoretical Medicine and Bioethics* 38: 367–385. https://doi.org/10.1007/s11017-017-9418-8.

Godfrey-Smith, Peter. 1994. A modern history theory of functions. *Noûs* 28: 344–362.

Hausman, Daniel M. 1995. The Impossibility of Interpersonal Utility Comparisons. *Mind* 104: 473–490.

Hausman, Daniel M. 2016. Health and Well-Being. In *The Routledge Companion to Philosophy of Medicine*, ed. Miriam Solomon, Jeremy R. Simon, and Harold Kincaid, 27–35. New York: Routledge. https://doi.org/10.4324/9781315720739-9.

Kingma, Elselijn. 2017. Disease as Scientific and as Value-Laden Concept. In *Handbook of the Philosophy of Medicine*, ed. Thomas Schramme and Steven Edwards, 45–63. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-8688-1_75.

Lemoine, Maël. 2013. Defining disease beyond conceptual analysis: an analysis of conceptual analysis in philosophy of medicine. *Theoretical Medicine and Bioethics* 34. Springer Verlag: 309–325.

Lewens, Tim. 2015. *The Biological Foundations of Bioethics*. Oxford: Oxford University Press.

Melina, Vesanto, Winston Craig, and Susan Levin. 2016. Position of the Academy of Nutrition and Dietetics: Vegetarian Diets. *Journal of the Academy of Nutrition and Dietetics* 116. Elsevier: 1970–1980. https://doi.org/10.1016/j.jand.2016.09.025.

Millikan, Ruth G. 1989. In defense of proper functions. *Philosophy of Science* 56: 288–302.

Muckler, Dane, and James Stacey Taylor. 2020. The Irrelevance of Harm for a Theory of Disease. *The Journal of Medicine and Philosophy* 45. Oxford Academic: 332–349. https://doi.org/10.1093/jmp/jhaa007.

Neander, Karen. 1991. Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science* 58: 168–184.

Powell, Russell, and Eric Scarffe. 2019. Rethinking "Disease": a fresh diagnosis and a new philosophical treatment. *Journal of Medical Ethics* 45: 579–588. https://doi.org/10.1136/medethics-2019-105465.

Reznek, Lawrie. 1987. *The nature of disease*. London; New York: Routledge & Kegan Paul.

Rossi, Mauro. 2011. Transcendental arguments and interpersonal utility comparisons. *Economics and Philosophy* 27: 273–295. https://doi.org/10.1017/S0266267111000216.

Schramme, Thomas. 2002. Is it bad to have a disease. In *Health and quality of life: philosophical, medical, and cultural aspects*, ed. Antje Gimmler, Christian Lenk, and Gerhard Aumüller, 61–67. Münster: Lit Verlag.

Schwartz, Peter H. 2014. Reframing the Disease Debate and Defending the Biostatistical Theory. *Journal of Medicine and Philosophy* 39: 572–589.

Sullivan, Gail. 2014. This bug's bite could turn you into a vegetarian. *Washington Post*, August 10.

Wakefield, Jerome C. 1992. The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist*: 373–388.

Wakefield, Jerome C. 1999. Mental disorder as a black box essentialist concept. *Journal of Abnormal Psychology* 108: 465–472. https://doi.org/10.1037/0021-843X.108.3.465.

Wakefield, Jerome C. 2005. On winking at the facts, and losing one's Hare: value pluralism and the harmful dysfunction analysis. *World psychiatry* 4: 88–89.

Wakefield, Jerome C. 2007. The concept of mental disorder: Diagnostic implications of the harmful dysfunction analysis. *World psychiatry* 6: 149–156.

Wakefield, Jerome C. 2013. Addiction, the Concept of Disorder, and Pathways to Harm: Comment on Levy. *Frontiers in Psychiatry* 4: 1–2. https://doi.org/10.3389/fpsyt.2013.00034.

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

Wakefield, Jerome C. 2021a. Can the Harmful Dysfunction Analysis Distinguish Problematic Normal Variation from Disorder? Reply to Andreas De Block and Jonathan Sholl. In *Defining mental disorder: Jerome Wakefield and his critics*, ed. Luc Faucher and Denis Forest, 511–536. Cambridge, Mass.: MIT Press.

Wakefield, Jerome C. 2021b. Must Social Values Play a Role in the Harm Component of the Harmful Dysfunction Analysis? Reply to Rachel Cooper. In *Defining mental disorder: Jerome Wakefield and his critics*, ed. Luc Faucher and Denis Forest, 553–575. Cambridge, Mass.: MIT Press.

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: 594–596. https://doi.org/10.1136/medethics-2019-105578.

Wakefield, Jerome C., and Jordan A. Conrad. 2020. Harm as a Necessary Component of the Concept of Medical Disorder: Reply to Muckler and Taylor. *Journal of Medicine and Philosophy* 45: 350–370.