Theoretical Medicine and Bioethics (2019) 40:1–19
https://doi.org/10.1007/s11017-019-09483-y

The harm of medical disorder as harm in the damage sense

David G. Limbaugh[1,2]

Published online: 2 March 2019

Abstract

Jerome Wakefield has argued that a disorder is a harmful dysfunction. This paper develops how Wakefield should construe *harmful* in his harmful dysfunction analysis (HDA). Recently, Neil Feit has argued that classic puzzles involved in analyzing harm render Wakefield's HDA better off without harm as a necessary condition. Whether or not one conceives of harm as comparative or non-comparative, the concern is that the HDA forces people to classify as mere dysfunction what they know to be a disorder. For instance, one can conceive of cases where simultaneous disorders prevent each other from being, in any traditional sense, actually harmful; in such cases, according to the HDA, neither would be a disorder. I argue that the sense of *harm* that Wakefield should employ in the HDA is dispositional, similar to the sense of *harm* used when describing a vile of poison: "Be careful! That's poison. It's harmful." I call this *harm in the damage sense*. Using this sense of *harm* enables the HDA to avoid Feit's arguments, and thus it should be preferred to other senses when analyzing harmful dysfunction.

Keywords Disorder · Disease · Harm · Systematic harm · Dispositions · Philosophy of psychiatry · Philosophy of medicine · Philosophy of biology · Philosophy of science

✉ David G. Limbaugh
dglimbau@buffalo.edu

[1] Department of Philosophy, State University of New York at Buffalo, Buffalo, NY, USA

[2] Romanell Center for Clinical Ethics and the Philosophy of Medicine, State University of New York at Buffalo, Buffalo, NY, USA

Introduction

Having a disorder can be characterized as being in a state of less than optimal health.[1] The flu, small pox, and the common cold are all disorders, but so are lupus, diabetes, and celiac. Similarly, on my view, bruised heels, paper cuts, and gun shot wounds also fall within this semantic range. Thus, an analysis of disorder should capture each of these, while excluding fever resulting from infection, hunger pangs during bouts without food, and a runny nose during a sinus infections. For each of these manifestations signals that the body is doing what it is supposed to be doing under the circumstances, which is to say, that the body is functioning properly. Proper function is nearly the antithesis of disorder.

Dysfunction, then, is at least a necessary condition for being a disorder. Whether dysfunction is also sufficient is currently under debate in philosophy of medicine. Notably, Christopher Boorse defends the position that disorders are merely dysfunctions, while Jerome Wakefield contends that disorders are harmful dysfunctions [1–4]. Below, I do not engage this well-trodden debate directly, but rather assume Wakefield's position at the outset so as to focus on his account's distinguishing feature—namely, the harmfulness of disorder.

The aim of this paper is to flesh out what Wakefield should mean by *harmful* when used as a necessary condition for disorder. While Wakefield has done much for philosophy of medicine in general and for philosophy of psychiatry in particular, his account of harm is largely underdeveloped. Furthermore, Neil Feit has argued that there is no plausible account of harm that can be paired with dysfunction in such a way that it results in a plausible account of disorder [5]. Thus, Feit argues, Wakefield should drop harm as a necessary condition.[2]

I argue below that Feit is wrong. The strategy is to develop an account of harm that fits Wakefield's purposes and then demonstrate how it avoids the counterexamples suggested by Feit. The account, in brief, is one of harm in the damage sense. When an artifact becomes damaged, sometimes the damage is readily apparent—like a knife that has broken from its handle. Other times, damage is revealed only when the artifact is used in a certain way. A knife whose blade has gone slightly dull is an example; the damage is apparent only when somebody goes to cut something requiring a very sharp blade; otherwise, the dullness may simply go unnoticed. Similarly, when one suffers from a certain dysfunction, one is sometimes damaged such that one is caused to either suffer intrinsic bads—things that are bad in and of themselves, like pain, depression, and the like—or become disposed to suffer intrinsic

[1] The relationship in natural language between the terms *disorder*, *disease*, *malady*, *illness*, *sickness*, and so forth is opaque. Hence here I use *disorder* as a catch-all technical term to capture the phenomenon that results in, or is, less than optimal health.

[2] Feit's objections are not uniquely tied to the harmful dysfunction analysis. Any account that tries to make use of the folk concept of harm in a philosophically rigorous way must overcome these objections. It is partially for this reason that Ben Bradley has suggested philosophers do away with *harm* in philosophical theorizing in favor of *bads* [6].

bads in circumstances where others would suffer not at all, much as someone with an enzyme deficiency might suffer pain when eating certain foods.[3]

The paper proceeds as follows. First, I lay out Wakefield's harmful dysfunction analysis. I then lay out Feit's objections to harm as a necessary condition for disorder. Building on this discussion, I introduce and develop a concept of harm in the damage sense. Finally, I demonstrate how the damage account avoids Feit's counterexamples and other objections to harm generally.

Disorder as harmful dysfunction

In this section, I lay out Wakefield's account. Wakefield considers a disorder to be a harmful dysfunction. This position forms what is called the harmful dysfunction analysis (HDA). By "disorder" Wakefield means roughly a state of lacking health. I am primarily concerned with the harm aspect of this account, though it will be helpful to elucidate what is meant by "dysfunction" as well. A dysfunction occurs when some internal natural (evolutionary) function is inhibited,[4] with internal natural functions understood as "effects that explain the existence and structure of naturally occurring physical and mental mechanisms" [1, p. 383]. He goes on:

> One can legitimately answer a question such as "Why do we have hearts?" or "Why do hearts exist?" with "Because hearts pump the blood." The effect of pumping the blood also enters into explanations of the detailed structure and activity of the heart. Thus, pumping the blood is a natural function of the heart. [1, p. 382]

I slightly alter Wakefield's terminology in this paper without altering the substance of his account. I do not call pumping blood the heart's *function*; I call it the heart's *functioning* [7]. Furthermore, a heart is functioning when its dispositions cause their effects, or manifest. Dispositions manifest, causing their effects, when they are in the right circumstances. One worries about the fragility of a fragile vase only when the vase is suitably struck or dropped, thus presenting the right circumstance for its fragile disposition to be triggered.

A function is a subclass of disposition. Functions are dispositions with a purpose. In the case of the HDA, they are dispositions with an evolutionarily selected purpose. So, rather than saying "Pumping blood is a natural function of the heart," I say, "Pumping blood is the natural functioning of the heart" or "The disposition to pump blood is the natural function of the heart." A dysfunction is present when a function is not able to manifest in those circumstances in which it would be triggered. Wakefield calls this an "unfulfilled function" [1, p. 381].

[3] I will at times use *suffer* in the sense of "have" when referring to a bad or harm. This not to imply that these bads or harms need be phenomenological in any way.

[4] I speak of evolutionary function merely to remain consistent with Wakefield. The account in this paper makes no commitment to any particular understanding of function.

Wakefield holds that the having of a dysfunction is not enough for the having of a disorder. The dysfunction must also be harmful.[5] This harm condition is thought to get the right result as regards the exclusion of benign dysfunctions from disorders. For example, having fused toes is not a disorder, even though it could be a dysfunction. I say "could be" given the difficulty of hashing out what functions there are and when they are impeded. If it is assumed that fused toes are a dysfunction, then there must be some evolutionarily selected-for function, which may no longer be relevant to our well-being, that is prevented from manifesting because of fused toes—thus rendering fused toes a dysfunction but harmless.

As a final point of clarification, it is important to note that there are two scopes of harm: *all things considered* and *pro tanto*. When an event is harmful all things considered, it has been assessed in terms of the whole event and all of its effects. When an event is pro tanto harmful, it has been assessed in terms of a specific aspect of the event, either in terms of some sub-event or in terms of only some of the event's effects.[6] For instance, if one takes surgery as a whole, then one takes it all things considered, but if when assessing surgery, one focuses on only the scalpel cutting the skin, then one is assessing the surgery pro tanto. In this way, surgery can be both all-things-considered beneficial and pro tanto harmful. Does the surgeon harm her patient when she makes an incision with a scalpel? Yes. Is the incision all-things-considered harmful? Likely not. Assuming the surgery is a success and that its benefits outweigh the harm of being cut into by a scalpel, it is clear that, though causing some harm, the incision is not harmful all things considered. It is merely pro tanto harmful.[7]

According to the HDA, a dysfunction need only be pro tanto harmful. This is important. Some disorders exclude other more harmful disorders. A classic example is cowpox, which, Wakefield says, "may be considered overall beneficial due to the protection it confers against smallpox" [4, pp. 668–669]. Thus, cowpox satisfies the harm conjunct of the HDA because it is pro tanto harmful, although it is not harmful all things considered.

With these preliminaries out of the way, a definition of a disorder according to the HDA can be given:

> **HDA.** A condition of person S is a disorder iff (1) it results from the inability of some internal mechanism to perform its natural function, and (2) it is pro tanto harmful to S.[8]

[5] This is *pace* Boorse, who champions the having of a disorder as "value-free," though even he admits of "a great variety of value-laden 'disease-plus' concepts" [3, p. 684]. Those adamant in their defense of Boorse's position should perhaps view Wakefield's harmful dysfunction as a value-laden subclass of Boorsian value-free pathology.

[6] Wakefield and Feit both seem to use *pro tanto* and *prima facie* interchangeably. I prefer to reserve the former for an ontological evaluation of harm and the latter for a merely epistemic evaluation.

[7] Theoretically, assuming there are atomic events, some events will have completely overlapping all-things-considered and pro tanto scopes of harm.

[8] This is the definition given by Feit [5, p. 368], which he formulates from Wakefield [1, p. 384]. I have added pro tanto to the definition, since it is implied by both Feit and Wakefield.

In the next section, I discuss issues raised by Feit in his argument against including harm in one's account of disorder.

The problem of harm

Here I discuss Feit's objections to the harm component of the HDA. There are traditionally two ways to understand when and to what degree someone is harmed: comparatively and non-comparatively. The comparative account is the more popular of the two and the prime candidate to accompany the HDA.[9] The comparative account counterfactually compares the actual world with how it would have been without the harmful state. It is concerned with the loss, or lacking, of goods as well as the gain, or having, of bads in comparison to how things would have been. The non-comparative account understands harm in terms of losing/gaining what is good/bad in and of itself—no comparison necessary. By "good" and "bad" I mean whatever accounts for positive and negative value in some theory of value. To avoid tying the discussion too closely to a particular value theory, I will say little about what goods there are, and I will do my best to stick to examples with forms that are easy to translate into the reader's preferred account. Nonetheless, it may end up that what is a disorder is partially determined by one's value theory.

I assume that there are two non-exclusive types of goods/bads—those that are good or bad in their own right (intrinsic goods/bads), and those that are good or bad because they are instrumental in the having of other goods or bads (extrinsic goods/bads). Pleasure is a common example of an intrinsic good, and pain a common example of an intrinsic bad. Pleasure is good no matter the circumstance; it is good *simpliciter*.[10] Money is an extrinsic good. It supplies spending power that can be used to acquire other goods. In this way, money must be in the right place at the right time. As such, it is merely instrumental in the having of other goods. Money is good because there are other intrinsic goods to be had.

Comparative account

I now turn to the comparative account and demonstrate how it fails as a complement to dysfunction for the sake of disorder. The comparative account of harm is as follows:

[9] Throughout this essay, the term *comparative* is elliptical for "counterfactual comparative." There are also temporal comparative accounts, but they handle the following objections no better than the counterfactual account, and face unique objections that an appeal to counterfactuals avoids; see [8, p. 368; 9, pp. 149–150].

[10] This would still be the case were the pleasure instrumental in the possession of some bads (the negative correlate of goods). There is no conceptual confusion in something's being an intrinsic good and an extrinsic bad, or vice versa.

> **Comparative harm.** A given actual condition, C, is harmful to person S iff S is in some state worse than what S would have been in if C had not obtained.

This is the general formula of a counterfactual comparative account. It draws no distinctions in the scope of harm (all things considered or pro tanto), and it is not committed to a particular theory of value. All it conveys is that figuring out whether C is harmful, and to what extent, requires a comparison between how well off one is in the world with C and how well off one is in the world without C, according to some scope of harm and value theory.

The HDA is interested only in the pro tanto scope of harm—that is, events that result in one's being worse off or not better off to at least some extent. By "some extent" I mean something similar to pro tanto as used in the context of harm. For example, assume that my receiving $50 in a thriving market is generally good because it provides me with more spending power. If the context of receiving the money were such that I am owed $100 but received only $50, then it would be right to say that having received $50 was good but only to some extent; there is more to consider—namely, that receiving $50 also partially constitutes the bad of being ripped off. The following presents an account of pro tanto counterfactual comparative harm:

> **Pro tanto comparative harm.** A given actual condition, C, is pro tanto harmful to person S iff S is, according to some state, worse off than what S would have been in if C had not obtained.

Here "some state" means any part of a circumstance. If *scraping my palms by jumping out of the path of an out-of-control car* is a circumstance, then scraping my palms, the car's being out of control, and my jumping are all parts of that circumstance. Though I am better off than I would have been had the car hit me (my ribs are not broken, for instance), I am still, according to some state (i.e., the state of scraping my palms), worse off than I would have been. After all, had I not jumped, then my palms would not be scraped!

Comparative accounts have notorious puzzles. The two types of troubling cases referred to in the literature are preemption cases and overdetermination cases.[11] I first discuss preemption—that is, when some condition preempts the occurrence of some other similar condition. Consider the following example situation:

> **Preemption Thugs.** I owe a bookie a large sum of money and forget to pay him. As a consequence, he sends Thug to break my legs, and Thug succeeds. The bookie knows that Thug is unreliable and so sends Brute to make sure Thug follows through. Had Thug not shown up to do the deed, then Brute would have broken my legs, in the same way at the same time, instead.

[11] For a sampling of the discussion surrounding such cases and proposed solutions, see [6, 10, 11].

According to the comparative account, Thug did me no harm. To be harmed is to be in some state worse than the state I would have been in if the situation had been avoided. The presence and reliability of Brute ensured that, were Thug to have not shown up, my legs would have broken anyway. If the goods I have in Thug world are compared with the goods I have in Brute world, there is no difference according to any state; and without a difference, there is no harm. However, undoubtedly, I was harmed when Thug broke my legs. Thus, the account delivers the wrong result.

The overdetermination puzzle is similar, but rather than one analogous event preempting another, both analogous events occur simultaneously:

> **Overdetermination Thugs.** This time Brute is overly zealous. Instead of waiting to see if Thug shows up, he rushes into the room to break my legs. Well, it just so happens that Thug also rushes into the room at the same time from another doorway. Neither one aware of the other, they both strike my legs simultaneously breaking them in the same spot at the same time.

The scenario here presents two conditions, the Thug-condition and the Brute-condition. Each of these is sufficient for breaking my legs. Thus, if the Thug-condition had not obtained, then the Brute-condition would have been enough to actualize a broken-legs world. The same holds for the Brute-condition: had Brute not acted, then Thug would have still broken my legs. According to the comparative account, neither the Brute-condition nor the Thug-condition harmed me. Only their conjunction was harmful.[12] When they are both present I end up in a broken-legs world, but if just one of them were absent, then I would still end up in a broken-legs world. The example assumes that the values of each of these worlds results in neither the Brute-condition nor the Thug-condition, alone, making me worse off than I would have been otherwise according to any state. Clearly something has gone wrong, since it is highly intuitive that each Thug and Brute, apart from their conjunction, harmed me when they simultaneously broke my legs.

As Feit explains, the form of these puzzles can be used to create counterexamples to the HDA:

> In a case in which pneumonia kills an old man and ends the suffering associated with a much worse disorder, it is plausible that pneumonia is a benefit. Here, Wakefield can say that pneumonia is a disorder if it puts the old man into bad states (chills, muscle aches, and the like) that he would not have been in without it, or if it prevents him from being in good states (the joy of seeing his spouse or grandchildren) that he would have been in without it. But it is possible to suppose that the pneumonia is not prima facie harmful in either of these ways. [5, p. 375]

[12] Feit's plural harm [12] can answer such challenges in the normative sphere. However, Feit himself has argued that when applied to disorder, plural harm elicits the wrong result. Hence I do not consider the view's merits in this paper.

Among the possibilities that Feit imagines is one in which pneumonia kills an old man who already has pneumonia-like suffering—which is to say, is worse off in pneumonia-like ways—but the suffering caused by this pneumonia-like condition is much worse. Thus, the suffering of the pneumonia-like condition effectively preempts the suffering that would have occurred had the pneumonia presented on its own. If it is stipulated that the pneumonia-like condition would have killed the man at virtually the same time had the pneumonia not been present, then there is no state in regard to which the man is worse off for having the pneumonia. Thus, according to the comparative account of harm, pneumonia may be neither harmful nor a disorder in the situation described above.

A more clinically tangible case where something like this could occur might involve atrial fibrillation (a-fib). A-fib occurs when the heart's atrium ceases to beat with a normal rhythm. Interestingly, there are a large number of documented cases in which the abnormal heart rhythm that is a-fib cannot be felt [13]. Mild a-fib can result in tiredness and shortness of breath. When a-fib is very severe, these side-effects are accompanied by decreased blood pressure, decreased oxygen to the brain, and the beginning of non-vital organ shutdown due to the body's reallocating resources to keep the brain, lungs, and kidneys from being damaged. Furthermore, if a-fib goes on long enough, then, because the abnormal rhythm causes blood to pool in the atrium, clots begin to form in the heart. When this occurs, what keeps those clots from (theoretically) exiting the right atrium, traveling through the right ventricle, and becoming clogged in the pulmonary artery is that the heart is not beating properly; it is not producing enough force to move the clots. Thus, in some instances, after clots form, a-fib can keep the victim from having a pulmonary embolism.

Here is the trouble. Were one to have a pulmonary embolism, then it is possible (and likely) that the resulting decrease in blood pressure, decrease in oxygen to the brain, and beginnings of non-vital organ failure would be more severe than those caused by a-fib. However, if the a-fib is what is keeping the victim from having a pulmonary embolism and, consequently, from having the worse symptoms, then the a-fib is not harmful according to the comparative account. Once the clots form in the atrium, the victim is not worse off with a-fib than she would be without it. Without this arrhythmia, the heart would pump the clots into the pulmonary artery causing a pulmonary embolism, making the victim worse off with the same, but more severe, symptoms (recall that there are cases where the victim cannot feel the discomfort of the arrhythmia). This means that before clots form, a-fib is a harmful dysfunction but afterward it is merely a dysfunction. Again, the comparative account yields the wrong result.

Non-comparative account

What about the non-comparative account? The objections posed in this section suggest it also fails. To be non-comparatively harmed is for some event to have caused a bad regardless of how the world would have been if the event had not occurred. Rather than harm amounting to the state of being worse off than one would have been otherwise, as in the comparative account, to be harmed is to be badly off

simpliciter. Pain is a common example of an event that harms without a need for comparison; other events might include mere physical discomfort or severe depression and loneliness.[13] The intuition is that, even if my severe depression were holding off some greater bad, my depression would be harmful in its own right. Because this determination of badness lacks the need for a comparison between states, the problems of preemption and overdetermination are avoided.

It is not controversial to assume that there are intrinsic bads, and what qualifies as an intrinsic bad depends on one's value theory. What is controversial is placing harm exhaustively within the scope of intrinsic bads. Imagine that my friend bakes me cookies as a surprise gift. My friend asks Thug to bring me the cookies, and unfortunately he eats them instead. I never find out, and neither my friend nor Thug ever tells me. In this case, it is plausible to think that I have suffered no intrinsic bads, and yet it is also plausible that I was harmed when Thug stole my cookies.

Of course, the advocate of non-comparative harm is within her rights to deny that I was harmed by Thug. Perhaps I am wrong to think I was harmed, and the non-comparative account can explain why. Whatever one thinks of this approach, it is much less clear that it works in the case of disorders. I present three of the cases mentioned by Feit [5] to make this point.

First, consider a case in which a young mother becomes infertile due to an infection. She and her partner had already decided they did not want to have any more kids before they heard the news. They consider themselves lucky that they had already made the tough decision to stop having kids, and they are thankful that this situation befell them and not some childless pair [5, p. 381]. One can imagine a case like this where the infertile mother suffers no intrinsic bads due to her infertility and thus suffers no intrinsic harms. But if there is no harm, then her infertility is not a disorder but a harmless dysfunction. The non-comparative account might shed light on the reasons for believing that there is no harm in this case, but if it does, then infertility would not be a disorder—a highly counterintuitive result.

Second are thought experiments where a person has a drastic loss of intelligence and yet suffers no intrinsic bads in the process.[14] Imagine a genius who enjoys her life as such. Tragically, Genius has a stroke and is reduced to someone of average intelligence. The stroke is a disorder, but intuitively so is the loss of intelligence. It is a stretch to assume that having average intelligence is intrinsically bad; as such, it is possible that Genius suffered no intrinsic bads, and no harm, by virtue of having become average. However, it is impossible for disorders to not be harmful. Thus, either the non-comparative account is false or Genius does not have a disorder [5, p. 381].

Third is an Epicurean worry about death. It is conceptually possible that death is the ultimate end and that we cease to exist after we die. If this is the case, then death cannot be intrinsically bad because there is no one left to be the subject of this bad. Thus, if some dysfunction were to kill someone immediately and painlessly—that

[13] For a defense of non-comparative accounts and more on lists of non-comparative harms, see [11; 14, p. 139; 15].

[14] This specific example is from Matthew Hanser [10, p. 432]. A similar, and well-known, example is found in Thomas Nagel [16, p. 77].

is, without intrinsic bads—then that dysfunction would cause intrinsic bads neither in one's dying nor in one's death. If disorder is to be the foil to health, then a dysfunction that kills is most certainly a disorder. However, on a non-comparative harm account, it is conceptually possible that no dysfunction that kills one immediately and painlessly is a disorder [5, p. 382].

Each of these counterexamples works by giving a case that presents a plausible disorder and showing why the dysfunction present does not meet the harm requirement on a non-comparative account. The previous section demonstrates a similar problem for the comparative account. A reasonable next step is to see if the harm of harmful dysfunction is disjunctive: comparative or non-comparative harm.

Unfortunately, this combination does not avoid situations where there is a disorder that manifests as a necessary lack of human flourishing. Feit gives the case of Evan to make this point [5, pp. 379–380]. Evan is someone who suffers from a disorder like Down syndrome. There are some who think that Derek Parfit was right when he supposed that the conditions entailed in our coming into existence are very particular [17]. It takes the right sperm and egg with the right DNA at the right time for "me" to result. In this way, Evan's Down syndrome might be a necessary condition of his existence. Were steps taken (e.g., manipulating the chromosomes of the egg) to make sure "Evan" was born without Down syndrome, then someone other than Evan might have been born in his place. However, if Evan's disorder is necessary, then he could not have been born without it, and thus he could not have been better off (every life is worth living compared to no life at all). This means that it is impossible for Down syndrome to be comparatively harmful. Furthermore, because Down syndrome does not necessarily cause pain, physical discomfort, depression, or symptoms along those lines, it is possible that Evan suffers no non-comparative harms either. Analyses based on comparative or non-comparative harm, then, both point to a determination in which Down syndrome is not a disorder on the HDA—which, again, is certainly the wrong result [5, p. 380].

The analyst is presented with several choices at this stage: (1) abandon harm as a necessary condition of disorder; (2) deny that the dysfunctions described above are in fact disorders; (3) appeal to some sense of *harm* that is conceptually distinct from those discussed above—one that is neither comparative nor non-comparative. The next section seeks to develop such a concept in keeping with this third option.

Harm as damage

This section develops my understanding of damage, demonstrating how this conceptualization serves to fulfill the harmfulness condition of the HDA. First, though, I must make some transitional remarks about harm. As discussed in the previous section, the comparative and non-comparative accounts both tap into something important about the different ways in which the term *harm* is used. Comparative harm treats harm as a negatively valued difference maker; as Bradley says, "It is based on an intuitively plausible idea: that harms make a difference, in a negative way, to the person harmed" [6, p. 397]. On the other hand, non-comparative harm zeroes in on harm as a bad state. Described by Elizabeth Harman, "Bad states are understood

as states that are in themselves bad, not bad because they are worse than the state the person would otherwise have been in" [14, p. 139]. In summary, the former is a matter of being worse off, while the latter is merely a matter of being badly off. As has been demonstrated by Feit, neither construal of harm is well suited to the HDA's sense of *harmful*, which in this context means more than causing one to be either worse off or in a bad state.

Rather, the sense of *harmful* in the HDA has a dispositional component absent in the above accounts, which allows it to avoid some, but not all, of the objections previously discussed (more on that below). It is not unlike the sense of *harmful* applied to poison in a vile, about which one might say, "Be careful—the stuff in that vile is very harmful." The poison in this instance is not presently causing harm, and yet it is still rightly called harmful. Likewise, disorders are harmful even while they lie dormant, causing no harm. This situation is typical of dispositions that manifest only when triggered. To say that a lump of salt is soluble is not to imply that the lump of salt is dissolving, but rather that it has a disposition to dissolve when in warm water. Likewise, a disorder is harmful not only in the sense that it is causing harm, but also in the sense that it would cause harm in certain circumstances. This sense of *harmful* is disjunctive; it can refer either to a thing's eventuation in actual harm or to a thing's potential to manifest harm in certain circumstances. Henceforth, I use *damage* to differentiate the dispositional (and disjunctive) sense of *harmful* from its typical philosophical usage.

The motivation to explain the harm of medical disorder using a dispositional concept like damage is best illustrated by example. Say Anna suffers from severe depression. She spends her days overwhelmed by anxious sadness. If it is assumed that depression is (or is caused by) a dysfunction and that the state of anxious sadness is harmful, then depression qualifies as a disorder. However, imagine that when Anna sleeps, her dysfunction remains but her occurrent states of anxiety and sadness disappear. Sleeping Anna has a dysfunction and is not suffering harm; yet one is still inclined to say she has a disorder. Here it is necessary to appeal to the fact that her dysfunction results in dispositions that do not always manifest harm but always dispose her to harm, rendering Anna damaged by her depression.

The benefit of an account like that laid out above is that it allows one to say that a person's reason for continuing treatment is to keep her disorder from harming her; were she cured, she could stop. Imagine someone who has diabetes and must go on a special diet to keep from having to take insulin. If such a person found herself loving her new diet with no desire to eat anything else, then, as long as the diet kept her diabetes under control, the diabetes would not result in any harm. However, in this case, one can still say that the diabetes is damaging. It results in a disposition that would manifest harm were one to eat food that others without said disposition could eat without the occurrence of harm.

As a disposition, damage is picked out by what it can manifest (what it can do) and its manifestation conditions (when it can do it) [18]. Broadly speaking, damage is a disposition that manifests harm in some situation. By "harm" I mean intrinsic bads of the sort associated with non-comparative harm. In the next section, I explain why, but for now I take this for granted. The sense of *damage* that is relevant to

disorder (and which is used for the remainder of this essay) is that of a disposition that manifests harm in societally relevant situations.

The meaning of societally relevant situations is in the spirit of Wakefield's "environmental circumstances" and "cultural standards" for deciding what harm matters for disorder [1, pp. 383–384]. According to Wakefield, "disorders are negative conditions that justify social concern" [1, p. 376]. Though concern is not a necessary condition, we rightfully care when we are disordered. Wakefield uses an example involving the removal of a single kidney to illustrate the efficacy of this condition:

> A dysfunction in one kidney often has no effect on the overall well-being of a person and so is not considered to be a disorder; physicians will remove a kidney from a live donor for transplant purposes with no sense that they are causing a disorder, even though people are certainly naturally designed to have two kidneys. [1, p. 384]

For my purposes, the phrase "often has no effect on the overall well-being of a person" should be understood as a statement about systematic harm. For an event to often have no effect on my well-being is for it to not harm me in typical situations—that is, situations that members of my reference class have a certain probability of encountering, given the laws of nature and the history of the world.[15] These are the societally relevant situations, and if some event would harm persons in enough of them, then there is cause for social concern.

Importantly, it is not a physician's sense of whether she is causing a disorder that matters in determining whether the damage caused will manifest in societally relevant situations. Nor would it matter if the kidney were removed against the will of the patient, say, by some highly skilled band of organ thieves.[16] All that matters is whether the patient (or victim) has some disposition that manifests intrinsic bads in societally relevant situations, which is an empirical question. In the kidney example, the assumption is that members of the patient's reference class are typically in situations where the having of one kidney does not manifest harm. Whether or not this is true is not determined by any one person, including the patient. Moreover, if the assumption is false, then Wakefield is wrong and the having of one kidney is in fact a disorder. Establishing the sense of typical that constitutes what is, or is not, a societally relevant situation is part of the project of pathologists, clinicians, and anthropologists. For now, the intuitive sense of typical as "normal" is sufficiently informative for the sake of formulating an operative notion of typical situations.[17]

I am now close to a reformulation of the HDA that analyzes disorder as a dysfunction resulting in dispositions that manifest harm in societally relevant situations.

[15] Here I follow Boorse in understanding reference class as "a natural class of organisms of uniform functional design; specifically, an age group of a sex of a species" [3, p. 684]. However, there may be other, more restricted meanings of *disorder* that rely on other determinations of reference class. For instance, members of the deaf community, when speaking in the context of deaf culture, may use the term *disorder* in a way that restricts a class to members of their community. See also [2].

[16] It is important not to confuse the physical state of having one kidney with the mental state of knowing that one's kidney was stolen. The former is a plausible dysfunction, while the latter is not.

[17] The verb *result* is intended to be read in a temporal or atemporal sense.

However, such a formulation would still fail to get the right result in all situations—specifically, it gets the wrong result in cases involving two disorders that are damaging separately but not together. For instance, imagine that dysfunction *x* causes sinus pressure from mucus buildup, and dysfunction *y* causes one's sinuses to dry out resulting in painful breaks in nasal lining. When these dysfunctions present together, one's nose has just the right amount of mucus and moisture. A good account should render both x and y disorders even in such cases.

An isolation clause is needed to ensure that harmful dispositions are always considered in their proper circumstances. To capture the plausibility of such a clause, one need only examine how someone would answer the question, "Why is a sinus infection damaging?" The best answer would undoubtedly begin with an account of what a sinus infection's resulting dispositions do in isolation from other dysfunctions. After all, the question is about the condition that is a sinus infection, not the condition "sinus infection and dysfunction *y*." Just because some dysfunction is not currently in a circumstance sufficient for its resultant dispositions to manifest harm does not mean that those dispositions would not still manifest harm in other circumstances. If those circumstances are societally relevant situations for individuals without any other dysfunctions, then those dispositions are damaging. Thus, when checking to see if some dysfunction is damaging, one has to look at the dysfunction in isolation.

Putting this all together, first one has a condition for something's being damage:

> **Damage.** Subject S is damaged iff S is suffering actual intrinsic bads, or S has some disposition that would manifest intrinsic bads, in societally relevant situations, in individuals without any other dysfunction.

Here is the HDA with "harm" switched out for "damage":

> **Revised HDA.** A condition of person S is a disorder iff (1) it results from the inability of some internal mechanism to perform its natural function, and (2) it results in damage.

In the next section, I use the revised HDA to revisit Feit's objection that the classic paradoxes of harm prevent the HDA from properly identifying disorders as such, and I push back on his suggestion that Wakefield would be better off identifying all dysfunctions as disorders.

Harm revisited

The damage account of harm resolves some of the objections lodged against the HDA because it maintains certain advantages over the comparative and non-comparative accounts, which are concerned with occurrent harm rather than dispositional harm. These advantages, which will emerge below, allow the cases from the third section to be addressed.

To begin, let me speak to why the damage relevant to disorder invokes non-comparative intrinsic bads, rather than the comparative harm of being worse off. At first blush, it would seem that preemption and overdetermination scenarios—such as the a-fib/pulmonary embolism case and "Overdetermination Thugs"—can be accounted for by virtue of the damage condition's isolation clause, which allows other competing dysfunctions to be screened off. Unfortunately, conceptual space is rarely so easily tidied. Though the isolation clause screens off rival dysfunctions, it does not account for preempting negative states that mask equally harmful evolutionary functions. Thus, were damage to manifest comparative harm, rather than non-comparative harm, then a general formula for generating counterexamples to the revised HDA could be given.

> **Comparative damage counterexample formula.** It is conceptually possible that some evolutionary function is such that it contributed to the reproduction and survival of our ancestors by manifesting discomfort of some sort. If some condition were to prevent this function from manifesting, then this condition would rightly be a dysfunction. Furthermore, if this dysfunction were to result in dispositions that manifest discomfort in societally relevant circumstances among individuals without any dysfunctions, then intuitively the dysfunction would also be a disorder. However, if the discomfort of this disorder is the same as that of the function it preempts, then individuals with the disorder would not be worse off than they would have been in societally relevant circumstances otherwise, and thus they would not be harmed or damaged.

This passage presents a case in abstraction that is similar to the preemption cases given in above. Even with revisions to the HDA, what is intuitively a disorder is rendered a mere dysfunction by the comparative account of harm. Now, I am not denying the possibility that the damage account can be further developed so as to rule out counterexamples to comparative harm. However, I suspect that Wakefield's meanings of both *function* and *dysfunction* would have to be altered also. On the other hand, non-comparative intrinsic bads easily explain preemption and overdetermination cases. Intrinsic bads concern only whether some event is by its nature a bad state. Thus, although in some preemption cases, a-fib may be keeping a pulmonary embolism at bay, a-fib is still actually resulting in intrinsic bads; coupled with the fact that a-fib is a dysfunction, this permits the conclusion that a-fib is a disorder in such cases.

I now turn to revisit Feit's cases, beginning with Evan. Evan necessarily has Down syndrome. Having Down syndrome is thought to result in a lack of flourishing but without any intrinsic bads (i.e., no pain or frustrated desires). Is Evan's Down syndrome damaging? The fact that Evan necessarily has Down syndrome does not prevent one from assessing Down syndrome according to the damage criterion. It serves to remember Charlie from Daniel Keyes' *Flowers for Algernon* [19]. Charlie, who suffers from phenylketonuria, is initially content with having an IQ of 68 and working a menial job. However, after an experimental procedure allows him to temporarily experience life and society as a genius, returning to subpar intelligence causes him to exchange his contentment for lament. Keyes' story demonstrates that

one can consider conceptually what it would be like for someone with no dysfunction to take on some of the dispositions resultant from the likes of phenylketonuria or Down syndrome. The upshot is not that Down syndrome might cause intrinsic bads. Rather, this exercise motivates the intuition that if the dysfunction of Down syndrome were taken on in societally relevant situations by someone with no other dysfunctions, then the resultant dispositions would manifest intrinsic bads similar to those experienced by Charlie. Thus, although Evan's Down syndrome is necessary for his existence and effectively not bad for him, there is no trouble conceiving of Down syndrome as a damaging dysfunction.

The case of Genius is similar. Genius is someone whose brain injury has caused a dysfunction such that her intelligence drops to average level without resulting in any intrinsic bads. Resolving this case turns on the observation that the question of whether Genius, in this single instance, suffers intrinsic bads from her injury is ultimately irrelevant. Rather, the relevant question is whether or not this drop in intelligence results in dispositions that, according to the reference class, would manifest harm in societally relevant circumstances. If it would, then Genius has a damaging disorder. While intuitively it may seem that a sudden drop in intelligence would be damaging, this remains an empirical question, which is difficult to answer from the armchair. Furthermore, I suspect that were one convinced that a sudden drop in intelligence would not result in damaging dispositions, then the motivation to call it a disorder would decrease. By my lights, what drives the intuition that a sudden loss of intelligence is a disorder is the belief that such a loss would generally result in a number of bad states. Were one to discover that this is not the case, then the desire to call Genius's condition a disorder, rather than a dysfunction, would dissolve. To say otherwise is to suggest that merely being of average intelligence is in and of itself unhealthy.[18]

The young mother objection turns on there being some case in which a young mother's being infertile does not cause her intrinsic bads. According to the original formulation of the HDA, the young mother who does not desire to have more children would not have a disorder in the form of sterility because her dysfunction is not harmful. The same conclusion is not reached using the revised HDA. As has already been demonstrated, a single case is not enough to show that something is not a damaging dysfunction. Rather, it must be shown for cases of infertility usually according to the mother's reference class. To this end, one has to examine the nature of the dispositions involved in infertility and their manifestations in societally relevant circumstances. It is relatively clear that in present society most cases of infertility will result in damaging dispositions. For one, most members of the young mother's reference class would not rejoice at being infertile but would instead lament this news. Furthermore, many of them would discover their infertility only after a long period of trying to conceive, thereby manifesting severe disappointment. The reactions of

[18] It is worth noting that there may be cases where social factors obscure whether something is a disorder or not. Say the makeup of society results in an equal split in regard to whether or not a dysfunction's dispositions manifest intrinsic bads, according to the reference class, in societally relevant situations. In this case, it might be unclear whether or not a dysfunction is a disorder.

the members' communities and broader societies are also relevant. It is not unlikely that feelings of isolation and loneliness would develop in the afflicted by virtue of an unfortunate societal sentiment that there is "just something wrong" with infertile young women.

Note that this last point is not to endorse a subjectivist view on the question of whether a young mother's infertility is a disorder. Mere opinion cannot make some dysfunction a disorder; to be a disorder, a dysfunction must result in a disposition to be objectively harmed. If society's opinions (positive or negative) result in behavior that is harmful to infertile young women, then they can be enough for an infertility-dysfunction to result in dispositions that are damaging. Thus, though societal opinion can play a role in turning a mere dysfunction into a disorder, what does the work is the objective harm incurred by the young mother, not popular opinion.

Notably, if society were to shift away from its emphasis on fertility as an important feature of being a young woman, then infertility might cease to be a disorder. If, for instance, the only intrinsic bads manifested by the dispositions of infertility were feelings of depression, disappointment, isolation, and so forth, then a large enough shift in the attitude of society—how infertile young women are treated and viewed, as well as how they view themselves—might result in the reclassification of infertility as a mere dysfunction. Of course, depending on one's value theory, the intrinsic bads caused by infertility may extend beyond those of mental experiences (see [20]).

The general idea here can be applied to any dysfunction that might carry a social stigma. Dermatological conditions are ripe for this sort of analysis. There are variants of conditions like acrochordons, cherry angiomas, melasma, psoriasis, and vitiligo that are merely cosmetic, but in a society that values certain appearances can result in harm through stigma and discrimination. To put it informally, the reason we seek treatment for a number of dermatological conditions is because we do not like the way they make us look and (unfortunately) neither does our society; this is enough for such conditions to be rightly considered disorders.

Finally, there is the Epicurean death objection. How can a dysfunction be harmful if it results in a painless—that is, intrinsic bad-less—annihilating death? In this case, there is no subject to have even suffered potential harm. If this objection is successful, then one's value theory must render it the case that the event of death itself, which results from a dysfunction, is not a case of damage. Here I give two examples of value theories according to which death causes intrinsic harm, and thus is a case of damage (albeit only for an instant). The goal here is to show that there is no conceptual problem with death's being a disorder according to the revised HDA, thereby allowing the status of death to be determined by one's value theory. I am sympathetic to the intuition that even painless deaths must be disorders; however, if death is not intrinsically harmful, then my personal sympathies at least are greatly reduced.

The first value theory takes the eradication of one's capacity for autonomy, which many would assume death involves, as an intrinsic bad for the subject. The accounts of value and harm put forth by Seana Shiffrin and Matthew Hanser engender support for such a view [10, 11]. According to Hanser, "'Goods' are not states or conditions that it is good to be in. Rather, they are things that it is good to have. And 'basic' goods are, roughly speaking, those the possession of which makes possible

the achievement of a wide variety of the potential components of a reasonably happy life" [10, p. 440]. It is easy to see how a capacity for autonomy (or autonomous action) should be considered such a good, according to Hanser. Without autonomy, we cannot accomplish our projects, fulfill our self-conscious desires, or be considered ends in and of ourselves. Shiffrin bolsters this point in arguing for her own account, which takes a capacity for autonomy to be the intrinsic good that is protected by autonomy rights:

> Its roots [the value of autonomy] lie foremost in a sense of the significance of the separateness of persons and the value of their separateness. Autonomy rights respect that an individual's will is distinct and separate from others' by respecting a domain in which that will is sovereign. They respect that the individual has a special, intimate relation to her mind, body, experience, and environment that she must especially endure, rendering it fitting that she and not others (whose relation and exposure to these experiences and conditions is more distant and indirect) exerts control over it—whether that control represents an especially substantive expression of character, rationality, or other intellectual virtues. [11, p. 382]

If death means annihilation, then death necessarily destroys one's capacity for autonomy, thereby undoing any ability to acquire a reasonably happy life or to be sovereign over one's mind and body. Such destruction is an intrinsic bad.

The second value theory that regards death as an intrinsic bad appeals to the non-subjective interests of the subject. For instance, David Hershenov thinks it is a mistake to confuse "something being in an individual's interest" with "that individual taking an interest in something" [21, p. 136]. In short, it might be in an organism's interest to be caught up in a life even when that organism cannot take an interest in its life. Thus, if some dysfunction causes death in an organism, then that organism's interests have been frustrated. The point is not that the organism is worse off than it would have been, but rather that the organism is simply not as it should be. This disparity between the organism and how it should be would constitute an intrinsic bad, thereby making any dysfunction that results in death a disorder.[19]

If either of the above accounts is correct, then a deadly dysfunction is necessarily damaging; and thus, if death is a dysfunction, then it is a disorder. Of course, the relationship between death and harm is cavernous, and not to be fully fleshed out here. The problem of death is difficult for any account of harm to accommodate. Even the comparative account, which would seem the most suited to handle an extrinsic loss of goods such as death, faces puzzles (see [22, 23]). As such, all I have tried to show in this case is that there is no conceptual trouble for an account in which deadly dysfunctions are disorders, though it might take the right value theory to work out the details.

[19] What about death resulting in an afterlife of eternal bliss? Whether death is a disorder will depend on the details of how one gets to there. As long as the animal is destroyed in the transition, Hershenov's account can accommodate there being intrinsic bads. If the person is destroyed and later begins to exist again, then perhaps appealing to a temporary loss of autonomy would be sufficient.

Conclusion

In this paper, I have argued that a dispositional notion of harm, which I have referred to as *damage*, is important to understanding what differentiates a dysfunction from a disorder. Incorporating such an account of harm helps to solve many of the puzzles faced by those inclined to analyze disorders in terms of harmful dysfunctions. I make no claim as to whether damage is helpful in moral theorizing or whether it can be used outside of philosophy of medicine at all. However, it is clear that this concept of harm as damage tracks an important natural usage of the term *harm*, and I conclude that this usage is what should be applied when considering Wakefield's harmful dysfunction analysis.

Acknowledgements For helpful feedback on earlier versions of this paper, I am very grateful to James Delany, Neil Feit, David Hershenov, Robert Kelly, Steven Kershnar, Danielle Z. Limbaugh, Kathryn E. Limbaugh, Sean Marzolf, Travis Timmerman, Neil E. Williams, and the fellows at the Romanell Center for Clinical Ethics and the Philosophy of Medicine. A special thanks to the anonymous referees at this journal for their helpful and thorough comments.

References

1. Wakefield, Jerome C. 1992. The concept of mental disorder: On the boundary between biological facts and social values. *American Psychologist* 47: 373–388.
2. Boorse, Christopher. 1977. Health as a theoretical concept. *Philosophy of Science* 44: 542–573.
3. Boorse, Christopher. 2014. A second rebuttal on health. *Journal of Medicine and Philosophy* 39: 683–724.
4. 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.
5. Feit, Neil. 2017. Harm and the concept of medical disorder. *Theoretical Medicine and Bioethics* 38: 367–385.
6. Bradley, Ben. 2012. Doing away with harm. *Philosophy and Phenomenological Research* 85: 390–412.
7. Spear, Andrew D., Werner Ceusters, and Barry Smith. 2016. Functions in basic formal ontology. *Applied Ontology* 11: 103–128.
8. Holtug, Nils. 2002. The harm principle. *Ethical Theory and Moral Practice* 5: 357–389.
9. Norcross, Alistair. 2003. Harming in context. *Philosophical Studies* 123: 149–173.
10. Hanser, Matthew. 2008. The metaphysics of harm. *Philosophy and Phenomenological Research* 77: 421–450.
11. Shiffrin, Seana Valentine. 2012. Harm and its moral significance. *Legal Theory* 18: 357–398.
12. Feit, Neil. 2015. Plural harm. *Philosophy and Phenomenological Research* 90: 361–388.
13. Page, Richard L., Thomas W. Tilsch, Stuart J. Connolly, Daniel J. Schnell, Stephen R. Marcello, William E. Wilkinson, Edward L.C. Pritchett, and Azimilide Supraventricular Arrhythmia Program Investigators. 2003. Asymptomatic or "silent" atrial fibrillation: frequency in untreated patients and patients receiving azimilide. *Circulation* 107: 1141–1145.
14. Harman, Elizabeth. 2009. Harming as causing harm. In *Harming future persons: Ethics, genetics and the nonidentity problem*, ed. Melinda A. Roberts and David T. Wasserman, 137–154. Dordrecht: Springer.
15. Shiffrin, Seana Valentine. 1999. Wrongful life, procreative responsibility, and the significance of harm. *Legal Theory* 5: 117–148.
16. Nagel, Thomas. 1970. Death. *Philosophy and Public Affairs* 4: 73–80.
17. Parfit, Derek. 1984. *Reasons and persons*. Oxford: Oxford University Press.
18. Vetter, Barbara. 2015. *Potentiality*. New York: Oxford University Press.

19. Keyes, Daniel. 1959. *Flowers for Algernon*. San Diego: Harcourt.
20. Hursthouse, Rosalind. 1991. Virtue theory and abortion. *Philosophy and Public Affairs* 20: 223–246.
21. Hershenov, David. 2016. Death, dignity, and moral status. *University Faculty for Life and Learning* 26: 119–142.
22. Bradley, Ben. 2009. *Well-being and death*. New York: Oxford University Press.
23. Timmerman, Travis. 2016. Your death might be the worst thing ever to happen to you (but maybe you shouldn't care). *Canadian Journal of Philosophy* 46: 18–37.