

Wakefield on Harm, Health, and Homosexuality

**Christopher Boorse
University of Delaware
November 2021**

Draft only – not for quotation

For nearly thirty years, Jerome Wakefield has defended his “harmful-dysfunction analysis” (HDA) of medical disorder. In current summary, it

claims that “disorder” refers to “harmful dysfunction,” where dysfunction is the failure of some feature to perform a natural function for which it is biologically designed by evolutionary processes and harm is judged in accordance with social values. (202x, 27).

Like my own “biostatistical” view, Wakefield has always been clear that his analysis aims to cover “every pathological condition” (1992b, 233) -- including all diseases, which, like me, he views as a subclass of disorders (2000, 20). He also describes his HDA as a hybrid of facts and values: facts about function, values about harm. Although he offered a detailed theory of biological function (1999) soon after his original HDA paper, until fairly recently he had said little to clarify his harm component. But recently he has given a detailed explanation of harm as well (2013, 2014, 2021fg; Wakefield and Conrad 2020).

In the light of this new work, I shall defend three theses. First, the HDA would be better off if Wakefield simply deleted all his remarks on social values. Besides being unnecessary to the HDA, those that are nontrivial rest either on a false environmental relativity of disease, a false cultural-relativist metaethics, or an unacceptable medical ethics. Second, even after this deletion of social values, the HDA fails to match medical diagnosis, since it excludes all subclinical disease, as well as many harmless clinical diseases, and it can also find no harm in either essential pathology or the diseases of nonsentient organisms. Third, contrary to Wakefield’s passionate desire, his newly clarified harm analysis leads straight to the conclusion that homosexuality is harmful, and so a disorder if it is a dysfunction.

Two points about the HDA should perhaps be noted at the outset. First, as his critique of me (2014) stresses, Wakefield insists that both clauses are necessary. He thinks that “separating these two components of disorder judgments” was “the HDA’s essential contribution” (2021f, 512), which accounts for much of its advantage over my own view. As we shall see, however, in many ways, including on the most popular explanations of harm to nonsentient beings, the harm clause threatens to collapse into the dysfunction clause, as DeBlock and

Sholl (2021) argue for different reasons. Second, we must bear in mind that the harm concept Wakefield intends is in the welfare family: harm is a kind of defect in well-being (1992a, 384; Wakefield and Conrad 2020, 352). Undeniably, the term ‘harm’, as a synonym of ‘damage’ or ‘injury’, is sometimes applied to inanimate objects incapable of welfare, such as houses or artworks. Usually when it is so applied, that is because such damage reduces the object’s value to beings who, like us, are capable of welfare. And often that is because the injury interferes with the object’s function: a house is harmed by a hurricane when the damage reduces its usefulness as a human dwelling, which is a matter of dysfunction. But even if one supposes that artworks have intrinsic value beyond their benefit to human beings, we do not think a painting better or worse off from its own standpoint. By lacking the subjective perspective stressed by Sumner (1996), artworks and other artifacts have no individual well-being comparable to that of people and other sentient beings, and it is to this kind of welfare that Wakefield’s harm is an injury. And that is natural, since only that kind fits our intuition of medicine as fundamentally a practical discipline aiming to help patients.

I. Cultural Relativity of Harm

From the beginning, Wakefield has explicitly relativized harm in the HDA to the social environment (1992a, 385; 2021f, 511; Wakefield and Conrad 2020, 351). A few of his examples are physical, or partly so. A person’s knockout mutation on one copy of the gene that produces C-III is beneficial in today’s society by reducing his triglyceride level (2021f, 519). Wakefield’s own caffeine addiction, he says, is harmless in his current milieu, where coffee is readily available (2017, 41). But most of his examples are psychological and depend on social values, illustrating what he calls

the inevitable degree of social relativity present in disorder status. If a failure of function has no impact on anything valued by a specific culture, it is not a disorder for that culture but merely a harmless dysfunction or anomaly. The HDA allows for an appropriate degree of such cultural relativity. (2021g, 557-8).

Thus, in the mental realm, Wakefield mentions the examples of a brain defect in preliterate culture that would block reading (2006, 164; 2013, 1); albinism in societies that do not care about skin color (2006, 164); and a high level of social

dependency in Japan vs. the U.S. (161). Another provocative case that he mentions is a charismatic but schizophrenic leader whose delusions are highly valued by his society (2021e, 463).

However, as these examples show, two ways in which one might think the disorder status of a trait to vary with social values must be clearly distinguished: by different effects, or by social evaluation of the same effects. As to the first, of course the effects of a trait can vary with environment, as when white moths become more visible against soot-darkened trees. In particular, in a social environment, a dysfunction can be harmful in one society but not another because of their different values. As Wakefield's central example of homosexuality well illustrates, a culture's approval or disapproval alone can make a huge difference to the harm of a personality trait. But the effect of social values can also be mediated by social institutions, as in the dyslexia example, where the harm is mainly caused not by mere disapproval, but by the dyslexic's lack of access to the important advantages of reading (2021f, 513-4). An intermediate case might be Wakefield's example of inadequacy in "social roles" (514-5). Wakefield explicitly rules out mere social condemnation as making a dysfunction a disorder, saying that the harm must be caused "directly" (513; Wakefield and Conrad 2020, 360), but he thinks the other two kinds of example illustrate genuine value-based harm.

However, such examples do nothing to show that disease status varies with culture. On the contrary, as physical cases illustrate, there is no such medical concept as "pathological in environment E." Conditions are normal or pathological *tout court*, not in a specific environment. What there is instead, first, is a concept of "pathogenic in environment E." Nordenfelt imagines a Swede in Africa whose white skin gets badly sunburned due to lack of melanin; Engelhardt imagines an African in Scandinavia whose black skin gives him rickets from lack of sunlight. But these examples confuse 'pathologic' with 'pathogenic'. In medical thought, neither white nor black skin is itself pathological; each is a normal polymorph, well-adapted to one environment in our species' range but not to another. As I have often noted before (1977, 548-9; 2011, 39; 2021, 33-35), adaptation and health are two different concepts, one environmentally relative, the other not. Conversely, there is also a medical concept of masked or compensated disease. Phenylketonuria or favism may have no ill effects in a dietary environment free of phenylalanine or fava beans; allergy to wasp venom may be harmless with no wasps around; diabetic symptoms may vanish with careful glucose monitoring and easy access to insulin; immune deficiencies may be harmless in a sterile bubble. Yet these underlying conditions remain pathological

-- as is Wakefield's addiction if, as he argues, it is a dysfunction. So his examples above, rather than being arguments for his harm clause, are actually evidence against it.

Nearly all writers, confusing health with adaptedness, like Wakefield allege a false environmental relativity of disease. But it is still surprising that he himself succumbs to this error, since he gets two related questions exactly right. One is my point above about normal variants' causing disease in unsuitable environments. He discusses this issue in reply to Garson's use of a recently popular example of "developmental mismatch." A crustacean, *Daphnia cucullata*, grows a tough, helmet-like head in waters full of predators; but in predator-free waters a helmet reduces mobility, putting a *Daphnia* that grew up in dangerous waters at a selective disadvantage in safe ones. Wakefield correctly matches medical thought by stating that such a *Daphnia*'s helmet, however maladapted it may be to its new environment, is no disorder: "developmental mismatches are not disorders" (2021d, 388). The other point he gets right is to firmly reject individual relativity of disease. Rachel Cooper (2005) holds that the same condition can be a disease for one person but not for another. DeBlock and Sholl are sympathetic, asking why the dyslexic who doesn't wish to read should be considered disordered (2021, 496). Wakefield gives two answers, of which the first is simply: "that's the way diagnosis works" (2021f, 512). But the same answer applies to environmental relativity: in ordinary medicine, as we saw, diagnosis is no more environmentally relative than it is individually so. If psychiatric diagnosis differs, that is just one way psychiatry does not fit the medical paradigm.

In any case, a second, quite different version of Wakefield's social-relativity claim has nothing to do with different effects. It is that the same trait T, with exactly the same effects, can be a disorder in culture C but not in culture C', because of their different values -- in particular, their different harm judgments. This is genuine "disorder for C" -- Wakefield's phrase above -- not "disorder in C." Again, this claim might be meant in two different ways. One is simply that C may judge T normal, while C' judges T a disorder. This thesis is not about what T is but about what it is considered to be, and as such is trivially true. The same point applies to dysfunction judgments: if two cultures disagree about what is a dysfunction, they will disagree about what is a disorder. The difference, for Wakefield, is that he believes there is objective fact about dysfunction, but not about harm (2021g, 558).

What Wakefield seems to be claiming, however, is this: T may not just be *considered* a disorder in C but not C', but may actually *be* a disorder in C but not

C', and neither C nor C' can properly dispute the other's diagnosis. This is clearly a metaethical view, cultural relativism, which is highly problematic and, in my opinion, thoroughly discredited. And it is very unattractive for psychiatry and for Wakefield himself.

To begin with, if harm is determined by social values, the HDA is not, as he claims, a hybrid of facts and values. Rather it conjoins two kinds of facts: facts about dysfunction, and facts about what a given society values. Whether American or Saudi Arabian society considers homosexuality or anosmia harmful is just as much a fact as whether they are biological dysfunctions. Worse yet, Wakefield's references to social values make his account vulnerable to objections parallel to all the standard ones against cultural relativism in ethics (*cf.* Rachels 2019, ch. 2). What are these? First, if construed as a meaning equivalence – 'X is harmful' means "Our culture judges X harmful" -- relativism starts an infinite regress. In judging that a condition is harmful, a society must really be judging that it judges it harmful, which is to judge that it judges that it judges it harmful, and so on, with no ultimate content to society's judgment at all. Besides this logical problem, if harm is a function of cultural judgment, then there is no room for one society to dispute the values of another, or for an individual to dispute the values of his own. Yet it makes perfect sense for Western culture to dispute the judgment of, say, Somali culture that young women are better off without labia and a clitoris, or for individual Somalis to reject this evaluation. And individuals may belong to a subculture of a culture – say, Somali Christians. When a subculture disagrees with the broader culture, which judgment applies?

These problems can easily arise within psychiatry. To use Wakefield's own examples, suppose "runaway slaves and Soviet dissidents" had a brain dysfunction that made them unusually "tolerant of oppression and more freedom-aspiring" (2013, 1). Then, it seems, given local judgments of harm, drapetomania and sluggish schizophrenia would have been genuine psychiatric disorders in the American South and the Soviet Union, respectively. Or take homosexuality, which has certainly been judged harmful by most human societies past and present, and therefore, if it involves dysfunction, would be, contrary to Wakefield, a genuine disorder in 19th-century America or 20th-century Russia.

Wakefield has recently tried to escape these conclusions by contrasting the "immediate reactions" (2013, 2) of a society with its own more considered judgments over time. The social views of harm relevant to disorder are to be

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 often-conflicting value commitments are most basic, how to adjudicate between competing values, whether some seeming values are really just rationalizations of unjust power or blind prejudice, and how changing circumstances should alter these judgments. (2019, 594).

But this move to “dynamic” values does little to remove the basic problem, since there is no guarantee that all societies are fundamentally opposed to, say, “unjust power” or “blind prejudice.” If all cultures had the same basic ethics, there would be no “minimal cultural relativity in the concept of disorder” (2006, 165) to begin with. This line also makes it impossible to know whether a given dysfunction is really a disorder today, since we cannot know what our own society will judge harmful (let alone just) tomorrow, or next century.

In any case, Wakefield and Conrad have recently explained that the HDA's reference to social values was based not on a relativist metaethics, but on the fact that in medical practice, disorder status is not determined by the patient's or doctor's own “personal view of harm” (2019, 594). Regarding the latter, they write:

A Christian Science physician ... may truly believe that it is less harmful to die of infection than to be saved by an antibiotic and thereby violate God's will, but following that personal view of harm would be malpractice. Medicine is a socially sanctioned profession that carries with it a corresponding obligation to alleviate harm as judged by society, not as idiosyncratically judged by the physician. (2019, 594)

However one views this example – and it is hard to see how a Christian Scientist could be a physician at all -- clearly there are opposite ones as well. A Soviet physician should not treat a dissident for sluggish schizophrenia, even on Wakefield's hypothesis of a dysfunction. An Eritrean physician, or a follower of Isaac Baker Brown, should not treat an adolescent girl for her orgasms, even if North-African culture or Victorian English culture thinks them both dysfunctional and harmful; he ought to know better. Ethical physicians should give no treatment that they ought to know is harmful – for example, “self-demand” amputation of a psychotic's healthy legs -- even if their medical society approves it, so they have

no fear of prosecution.

Fortunately for Wakefield, these are issues of medical ethics or law that he has no need to resolve. Whose values legitimate what treatment has nothing to do with the meaning of ‘disorder’. Obviously, if disorder status involves a value judgment, then an individual’s judgment of disorder will depend on his own values, and a society’s judgment of disorder will depend on its values. That is automatic from the inclusion of a value criterion. So, insofar as an official classification of disorders like the DSM represents a society’s opinion, its list of disorders will depend on that society’s judgment of harm. But issues about whose values are employed in judging disorder have nothing to do with the analysis of the concept. Wakefield should simply say that whoever judges disorder also judges harm. And having said that, there is no need to take any metaethical position, including Powell and Scarffe’s moral realism (2019). If disorder requires harm, then if harm judgments are objective and rationally justifiable, so are those of disorder, while if harmfulness is merely subjective, disorder is too.

Wakefield has come to realize this fact, though he still puts social values in every summary of his analysis and makes extensive remarks on basic metaethics (2021g, 555-62). He now calls his social-values test an “addendum” or “codicil” to the basic HDA, analogous to his specific analysis of biological function.¹ “[T]he social values addendum is not strictly part of the concept of disorder, and an alternative theory of human harm would be possible” (555). This is progress; but it would be even better for him simply to delete all his remarks on cultural relativity of disorder. As we saw, where they are not trivial, they rest either on false environmental relativity of disease, a false metaethics, or an unacceptable medical ethics. To delete all these ideas would merely strengthen his view without loss.

II. Inadequacy of the Clarified Harm Test to Biomedical Classification

A. Disease types, tokens, and harm. Any disease type is instantiated by its cases (what philosophers call “tokens”) in individual patients. Obviously, not every case of disease is a net harm to the individual: cowpox may prevent smallpox, asthma may excuse a young man from the draft. Writers like Reznek (1987, 161) and Cooper (2005, 26) deny that such harmless tokens are cases of

¹The analogy may be imperfect, however, in that Wakefield regards at least the etiological aspect of his theory of function as a matter of conceptual analysis.

disease at all. But that is not the medical view; in medicine, obviously, a wound is pathological even if medical leave for it saves your life in wartime. So it is to Wakefield's credit that he does not take this line.

A great many possible ways can be imagined to find an acceptably loose connection between disease types and harmful tokens thereof. One might propose that only most cases of the disease must be harmful, or only some of them, or one might say a disease must be often or typically harmful or harmful in normal circumstances. One might distinguish *prima facie* harm from net harm, or direct from indirect harm, or intrinsic from extrinsic harm. Or one might adopt some form of Spitzer's formula: that the condition in its "fully developed or extreme" (Spitzer and Endicott 1978, 19) or "full-blown" (Spitzer and Wilson 1975, 829) form must "regularly and intrinsically" (*ibid.*) cause harms like distress or disability. In the end, Wakefield and Conrad (2020) make some use of all of these options.

As to the last, though Wakefield and Conrad give Spitzer's ideas a lengthy and highly favorable mention, they seem clearly unsuitable for the HDA.² It is strange to classify conditions only by their extreme versions, since many normal conditions would be harmful in extreme form, as Cyrano de Bergerac, or anyone else deformed by monstrous organs, would testify. Spitzer might protest that a normal nose is already fully developed, and Wakefield can, of course, object that a normal nose involves no dysfunction. The serious problem is that every dysfunction is harmful in extreme or fully developed form, making the harm clause superfluous. This is true almost by definition on my analysis (1976, 1977) of a part's normal biological function as its species-typical contribution to individual survival or reproduction. Because of massive redundancy, a few dysfunctional skin or liver cells may have no appreciable effect on the organism. But if all the cells had the same dysfunction – if it reached "full development" -- organ failure would result, injuring fitness. On Wakefield's analysis of function, an effect favored by evolution in a past environment may have no survival benefit

²Perhaps contrary to Wakefield and Conrad's view, however great Spitzer's influence on DSM-III, he cannot be taken as an authority on the disease concept, since DSM is itself perennially controversial. Besides his unsatisfactory clause on "full development," by his requirement of a "call to action" Spitzer is also among those writers who commit the therapeutic-imperative fallacy: to define disease in terms of a need for treatment. Briefly, this is fallacious because there is often no present treatment for diseases, while to say a condition needs ideal medical treatment is no more than to say it is bad and needs the kind of treatment appropriate for diseases – that is, for biological dysfunctions. For more discussion, see Boorse (2021).

today, in which case its absence today is dysfunctional but harmless. To me, however, that is just one objection to selected-effects theories of function.

Besides Spitzer's unsatisfactory formula, in the most intensive clarification to date of the HDA's harm clause, Wakefield and Conrad (2020) settle on the following: a disease type is a dysfunction that, dispositionally and typically (354-7), causes significant direct *prima facie* harm (360) "under some range of standard circumstances" (357). Unfortunately, despite this definition's great complexity, it is still wholly unsatisfactory. As I have argued in detail elsewhere (Under review), early subclinical stages of diseases like emphysema, coronary atherosclerosis, and autoimmune insulinitis have no harmful effects on the organism's gross output. Other conditions, like carcinoids and slow-growing prostate cancer, openly violate Wakefield and Conrad's "typicality qualifier" (356): most cases never cause harm to the organism. There are also many types of skin pathology, such as warts and hemangiomas, that ICD-10 classifies as diseases and dermatology books explicitly call harmless. Remarkably, Wakefield and Conrad's solution to this problem is apparently just to reject the whole category of subclinical disease. "Here, too," they write of viral infections, "disease is distinguished from nondisease by the presence or absence of harmful symptoms."³ But subclinical disease either is, or is inseparable from, the most basic idea of scientific medicine, and has been so since the advent of Virchow's cellular pathology 160 years ago. It is hard to imagine a larger change in medicine's conceptual scheme than to claim that disease states with no symptoms or signs are not really diseases at all.

Two other general objections to a harm clause still remain: the problem of essential pathology, and the problem of harm to nonsentient organisms. Since Feit (2017, 378-82) has already treated the first at some length, I can be brief.

B. Essential pathology. Two widely-held philosophical views together seem to block any harm requirement on disease: genetic essentialism and the counterfactual analysis of harm. According to genetic essentialism, one type of

³2020, 358. By "too," the writers mean "just as we saw for bacterial infections." But there are also subclinical bacterial infections described as disorders, such as tinea versicolor or tinea pedis. Interestingly, in this paper the authors say nothing of Wakefield's thesis in (2014) that such conditions are not diseases of the person, only of his body parts, and only "biological pathology," not medical.

origin essentialism, some or all features of an organism's genotype⁴ are essential to its identity, in that any organism without them would have been a different organism. If the whole genotype is essential, then every genetic disease, such as phenylketonuria, Tay-Sachs, and Huntington's chorea, is inseparable from a person's identity. Even if only broader features of the genotype are essential, like chromosome number, diseases such as Down's syndrome (trisomy 21) or Turner's (XO) will still be essential. But by the counterfactual analysis of harm (Feinberg 1988), disease can harm a person only by making him worse off than he would have been without it. Without his essential diseases, he would not have existed at all. Even if one argues (Feit 2017, 379) that a person is worse off with a life of negative well-being than with no life, neither Down's nor Turner's patients seem to fit that description; the former, at least, are typically very happy.

Defenders of a harm requirement might take various paths to resist this objection. They might just reject the fairly small set of diseases based on gross genetic defects as genuine diseases. Or they could reject genetic essentialism, as Cooper (2015) argues for doing on other grounds. They could exchange the counterfactual analysis of harm for a noncomparative theory based on intrinsic evils – or, more plausibly, change the harm requirement to one of some other concept of the welfare family lacking the counterfactual feature. However, Feit (2017, 380-82) argues that this last move must fail because disease need not include anything intrinsically bad.

Since the essential-pathology objection may involve a fairly small group of diseases, and some routes to avoid it are available, perhaps it is less serious than a more general problem: how to explain what harm to nonsentient organisms might be, if not dysfunction.

C. *Harm to nonsentient organisms.* Sedgwick (1973, 30-31) once claimed that the health concept applies only to human beings and their pets and livestock. Other writers, like Fulford (1989, 122-3) and Nordenfelt, have agreed that at least it cannot apply literally to plants and lower animals [PLA], but only in an

⁴In its original form, Kripke's doctrine was that each of us necessarily originated from a particular sperm and egg, in which case genetic essentialism is just one version of material-origin essentialism. In the text I use a variant based on genotype, which avoids the objection that one's original sperm and egg could have had different genes. All these ideas have their own recent origin in Kripke (1980); their application to disease grows out of Parfit's nonidentity problem (1984).

analogical or “parasitic” sense.⁵ But these views contradict settled scientific practice. As I have shown elsewhere in detail (2014, 696-9), to biologists the health concept is panbiotic: they apply it freely throughout the whole realm of life. We often see generalizations about disease in works on evolution or ecology, and whole journals are devoted to diseases of plants and invertebrate animals. For that matter, the basic task of describing species morphology presupposes a normality concept, since damaged specimens, like Kass’s half-eaten butterfly (1975, 13-4), are ignored, not seen as polymorphs. Wakefield, to his credit, is clear that the HDA is to apply to all organisms. Indeed, in one paper, he uses the examples of a mutant white moth in a sooty environment and a mutant bacterium oriented to the wrong geomagnetic pole (1992a, 385-6). One could hardly ask for a more panbiotic view of health than to admit diseased bacteria.⁶

But how can nonsentient organisms be harmed or benefited? Of the three main approaches to welfare theory for human beings,⁷ all seem to presuppose sentience. Obviously nonsentient beings cannot enjoy hedonist welfare, since pleasure and pain are conscious states. Desire-satisfaction theories cannot apply to beings with no desires; yet desire seems to require at least some psychology, of which plants presumably, and bacteria certainly, have none. As for objective-list theories of welfare, most of the intrinsic goods typically included in such lists – knowledge, friendship, aesthetic appreciation, and so on – also have no application to PLA, while the few that might apply – freedom, goal-achievement – can only be understood for PLA in terms of the goal-directed behavior that, I have argued (1976), is precisely the basis of function or dysfunction claims.

These considerations strongly suggest that Singer (1994, 220), Sumner (1996), and others are right: life itself is of no value to the nonsentient organism. Rather, all the things that make life worth living require consciousness. Living things do share distinctive characteristics, such as growth, reproduction, and functional design aimed at certain goals. But none of these properties is unique to

⁵Nordenfelt (1987, 104, 143). However, he later (2007) argued that plants and animals can have “holistic” health after all, based on their capacity for goal-directed activity.

⁶It seems natural to describe invasion by bacteriophages as an infectious disease of bacteria, though this usage does not seem to be common in microbiology.

⁷For brief summaries of the major alternatives, originally catalogued by Parfit (1984), see Crisp (2017) or Muckler and Taylor (2020). Longer discussions are Sumner (1996) and Fletcher (2016).

organisms or can support a genuine welfare concept. Put an iron bar and an orange in a moist environment, and rust grows on one and mold on the other. But to block its growth no more deprives the mold of anything valuable than it does the rust. There is no axiological difference *per se* between chemistry and biochemistry. And artifacts like pianos and cars also have functional parts, some even exhibiting Sommerhoff goal-directed behavior (1950, 1959), like robovacuums and robopets. Yet someone who has the same concern for his cyberdog as he would for a real dog shows a sentimentality as confused as it is alarming; a cyberpet's existence is of no benefit to it. But if life cannot be good, nor death bad, for a nonsentient organism, then neither can health be good nor disease bad for one. Thus, insofar as harm is injury to welfare, nonsentient organisms cannot be harmed. Yet they can certainly be diseased, as Wakefield freely grants.

Nevertheless, at least two groups of recent writers defend some kind of welfare for nonsentient organisms. One group is a line of "neo-Aristotelian" ethicists including Geach (1956), Anscombe (1958), von Wright (1963), Hursthouse (1999), Foot (2001), Thomson (2001), Nussbaum (2006), and Thompson (2008).⁸ In general, such writers seek an objective foundation for ethics in von Wright's Aristotelian doctrine (1963, 45) that every living thing has a good of its own. The good of an organism is seen as its flourishing as a good member of its kind, which in turn consists in its pursuit of natural functions serving natural goals. For example, Hursthouse evaluates an organism as good – *i.e.*, a good X – insofar as

(i) its parts, (ii) its operations, (iii), its actions, and (iv) its desires ... well serve (1) its individual survival, (2) the continuance of its species, (3) its characteristic freedom from pain and characteristic enjoyment, and (4) the good functioning of its social group – in the ways characteristic of the species. (1999, 202)

Here items (1), (2), and (4) might be taken to define a nonsentient being's welfare -- except that (2) and (4) can seemingly conflict with it. Both Hursthouse and Nussbaum discuss mother birds' species-typical behavior of risking their lives to distract predators from their nests. Hursthouse thinks that a bird who fails to do

⁸Brief surveys of this literature are Lutz (2018), pp. 10-13, Rice (2015), pp. 381-4, and Odenbaugh (2017), pp. 1033-37.

this is defective, even if it saves her own life. But Nussbaum says that “altruistic sacrifice for kin” can be part of an animal’s own good, a move that might preserve the view that being a good specimen of an X (Hursthouse 1999, 198) is good for X’s.

The second group of writers are environmental ethicists.⁹ In contrast with the typical view of animal ethicists that welfare requires sentience, many such writers attribute to nonsentient organisms “biological interests” (Varner 1998, 68) created by their teleological structure of functions and goals (*e.g.*, Taylor 1986, 121-3). For example, in Varner’s original scheme

[a]n individual A has an interest in X if and only if (1) A actually desires X, (2) A would desire X if A were adequately informed and impartial across phases of A’s life, or (3) X would fulfill some biological function of some organs or subsystem of A ...¹⁰

This disjunctive account allows plants and lower animals to have interests under (3). Like Wakefield, Varner uses a selected-effects view of function, creating a problem: selected functions can be bad for an organism in a changed environment (McShane 2019). Basl and Sandler (2013a, 2013b, 2019) offer a similar account, with the added twist of openly embracing welfare for teleological artifacts, thus accepting the *absurdum* in the objection that Dussault (2018, 190) calls *reductio ad artificium*. Unlike Varner, they do not commit to a selected-effects view of functions, but they still face the same problem of reproductive functions raised by Hursthouse and Nussbaum. On any view of biological function, the function of a bee’s sting is to protect her hive, which kills her in the process. But it is very implausible that one free-riding bee who fails to sting is worse off for her refusal (FitzPatrick 2000, 63-4).

In any case, there are at least two reasons why none of these accounts of nonsentient harm is likely to help Wakefield’s HDA. First, it seems clear that at least the neo-Aristotelians, and arguably the environmental ethicists too, construe the welfare of lower organisms simply as health itself – either the negative health of freedom from pathology, in being a good specimen, or perhaps positive health in unusual flourishing. For nonhuman animals, Hursthouse says our evaluations

⁹For summaries of the relevant literature, see McShane (2019), Dussault (2018), and chapter 3 of Varner (1998).

¹⁰(1998), 68. Later, however, Varner retracted this account (2003), 415-6.

“are all concerned with good *xs* as *healthy* specimens of their kinds” (1999, 206, italics original). Thomson says that “What is good for a plant is obviously what conduces to its health” (2001, 56-7). But if one conceptual element of harm is health, then to make harm an element of disease threatens the HDA with circularity.

A second problem is that, as we saw, nearly all these writers define welfare for nonsentient organisms precisely in terms of biological functions. That makes dysfunction automatically harmful, so that Wakefield’s harm clause becomes redundant for nonsentient organisms. Yet, as we saw, Wakefield holds that the HDA applies univocally to human beings, moths, and bacteria. Once again, then, the effort to save his harm clause from basic objections ends in its collapse into a dysfunction clause. So it seems Wakefield has not escaped Dussault’s conclusion:

[P]roponents of harm-requiring accounts of disorder seem to face a dilemma. They can either adopt a partly biofunction-based account of harm, whose plausibility is questionable in the first place [because of harmful functions], and which ... makes part-dysfunction sufficient for disorder. Or alternatively, they can adopt a sentience-based account of harm ..., and implausibly make their account of disorder inapplicable to nonsentient organisms. (2020, 15).

At best, on neo-Aristotelian or environmental-ethics accounts, it seems Wakefield will be driven to say that dysfunction – or some specification thereof, such as organism-level dysfunction¹¹ – is intrinsically bad for nonsentient beings, but only instrumentally bad for sentient ones. But this position is so contorted that it is a powerful motive to discard a harm requirement entirely.

III. Wakefield on Homosexuality

According to Wakefield, a major merit of the HDA is that it explains and justifies the American Psychiatric Association’s 1973 decision to normalize homosexuality *per se*. He writes:

Psychiatrists avoided the incendiary issue of whether homosexuality

¹¹This possible substitute for Wakefield’s harm criterion is considered by Dussault (2020) and by me (Under review).

is caused by a dysfunction and instead overrode the traditional reproductive-harm value claim, arguing that what really matters from a values perspective is capacity for loving human relationships. (2013, 2).

Spitzer's ... reasoning was ... that whether or not homosexuality is caused by a dysfunction ..., in the contemporary context of overpopulation and widespread birth control among heterosexuals, the highest generally accepted normative goal of sexual-love relationships is (or is increasingly becoming) not reproduction per se but mutual interpersonal and sexual satisfaction. Thus, except for social oppression, no harm as judged by evolving social values need occur to the homosexual individual based sheerly on the choice of a same-sex love object. (2014, 675-6)

Remarkably, Wakefield fails to notice that neither of the two arguments in these passages even appears to make sense. As to the main argument, there is no reason why only one value can "really matter"; one goal's being the "highest" does not mean that others vanish. Among positive values, even if $A > B$, $A + B > A$. Even if love is better than babies, love plus babies beats love alone. So all homosexual couples who wish to be parents still suffer a significant direct *prima facie* harm under standard circumstances by their inability to create biological children. What Wakefield needs is a trend toward viewing the capacity for blood offspring, not as of lower value, but as of no value at all; but of course no such trend exists. The second argument, suggested by the words 'need' and 'sheerly', is that homosexuality is not a disorder because it is not harmful to all homosexuals. But Wakefield, like everyone, concedes that not all cases of a disease need be harmful, as with cowpox in a smallpox epidemic and asthma in wartime.

Actually, one might imagine that the harm of mutual sterility¹² does afflict all homosexuals, on the grounds that, like other objective goods, the capacity to engender children is valuable "whether or not the subject desires" it (Muckler and Taylor 2020, 336). In apparent agreement, Wakefield and Conrad write:

¹²There is another obvious harm to homosexual parents that is rarely discussed: their children are deprived of crucial childhood experience of both sexes rather than only one. But since sex-specific psychology is controversial, I place no reliance on this point here.

Biologically normal-range physical and mental capabilities would certainly appear on any such list of objective goods ..., irrespective of whether the individual exercises such capacities. (2020, 354).

And regarding heterosexual infertility, Wakefield says that loss of a socially valued capacity is harmful regardless of individual people's desires.

Whether a condition is a disorder is not determined by how the diagnosed individual subjectively happens to feel about the condition's effects but by more "objective" standards determined by the culture's value system. Thus, for example, infertility at prime childbearing age is a disorder even if a patient has decided not to have children because ability to reproduce is generally considered a valuable capability in our society and deprivation of this ability is considered a *prima facie* harm irrespective of benefits that might accrue. (2021g, 557).

Noting this seemingly open contradiction between Wakefield's lines on heterosexual and homosexual infertility, Dussault (Unpublished) patiently explores various ways the two views might be harmonized, ultimately concluding that none of them succeeds. As I see it, Wakefield's only hope of escaping flagrant contradiction is to drop his objective-subjective contrast, relying instead on his "typicality" qualifier, plus the dubious empirical assumption that most premenopausal women do want more babies. That is the line he and Conrad eventually took in 2020:

[T]here are many obvious examples of disorders that do not harmfully impact an individual – for example, the impotent celibate, or the infertile woman who does not want children – but are considered disorders only because they are *typically* or *dispositionally* harmful (2020, 354)

Then Wakefield and Conrad's answer to homosexual couples' infertility will no doubt be a parallel claim: most homosexuals either do not want to be parents, or, if they do, do not care whether their children have both partners' genes.

Of course, one might think that a single effect, like infertility, cannot be the end of a harm analysis. Rather, the true issue should be whether homosexuality is

typically a net, not a mere *prima facie*, harm. Counterbalancing the joint infertility, homosexuality actually has several advantages. Same-sex couples avoid not only any need for contraception, but also the impossible problem of getting along with the opposite sex. Homosexual males mostly avoid the intense pressure heterosexual men feel from their female partners toward the disaster of fatherhood. On the other hand, there are other major drawbacks to homosexuality besides the inability to make babies. It is highly inconvenient to be restricted to ultimate sex acts using organs not built by evolution for that purpose, which therefore are awkward, inferior substitutes for ones that are. Also, homosexuals' minority status hugely restricts their partner pool. The vast majority of potential partners of the right sex are of the wrong sexual preference. It is an awful burden to know that roughly 95% of the people you are attracted to have no chance at all of being attracted to you.¹³ So it seems very unclear that, on the whole, the disadvantages of homosexuality are usually fully balanced by its benefits.

However, Wakefield and Conrad explicitly reject typical-net-harm analysis. "Neither the HDA nor medical judgment," they say, "prioritizes net harm over 'some significant harm'" (360).

A harmful dysfunction is a dysfunction that causes significant direct harm as one consequence, termed "*prima facie*" harm by Boorse and Wakefield ..., to be distinguished from net or on-balance or all-things-considered harm that takes into account the overall balance of all harms and benefits. (360)

In that case, the HDA is immediately forced to declare homosexuality a disorder if it is a dysfunction. For the two drawbacks just mentioned – organ mismatch and partner-pool restriction – affect all homosexuals. Each is not just a typical, but a universal, significant direct *prima facie* harm under standard circumstances.

As Wakefield emphasizes, the issue of homosexuality is of broad importance not just for the HDA, but for the legitimacy of DSM-III and its successors. Today, as in 1973, there is wide agreement that the APA vote to depathologize homosexuality did not result from any new scientific knowledge. Many early critics of the move called the change purely political: homosexuals,

¹³Estimates of the prevalence of homosexuality vary widely, depending partly on its definition. But for exclusive adult homosexuality, 5% looks more likely to be too high than too low.

they said, but not others on the traditional perversion list, had created a powerful lobby to change public attitudes. As a result, gay men and lesbians escaped from the pages of DSM, while fetishists, pedophiles, voyeurs, exhibitionists, sadists, masochists, bestialists, and necrophiles remained trapped therein. Both Spitzer's formulas and Wakefield's HDA purport to refute this charge of political nosology by offering a principled reason to declassify homosexuality, but not the other old perversions. So it is of great interest if neither effort succeeds -- Spitzer's account because it is a clearly inaccurate definition of medical disorder, and Wakefield's because, in its newly clarified form, it leads inexorably to the conclusion that homosexuality is a disorder if it is a dysfunction.

IV. Conclusion

In the first section, we saw that Wakefield can avoid all criticisms related to cultural relativity by detaching all his references to social values from his harmful-dysfunction analysis, then discarding his doctrines about them. But in our second section, we saw that the HDA must still fail to fit medical diagnosis, because of the vast number of harmless subclinical diseases (not to mention many clinical ones), and the impossibility of finding harm in essential pathology or nonsentient organisms. And in the third section, we saw that the HDA, far from explaining and justifying the APA's homosexuality decision, actually explodes its supposed rationale. In my view, then, Wakefield's recent clarifications of the harm clause are best described as the suicide of the harmful-dysfunction analysis.

BIBLIOGRAPHY

- Anscombe, G.E.M. (1958). "Modern moral philosophy." *Philosophy* 33:1-19.
- Basl, John and Sandler, Robert (2013). "The good of non-sentient entities: organisms, artifacts, and synthetic biology." *Studies in History and Philosophy of Biological and Biomedical Sciences* 44(4):697-705.
- Boorse, Christopher (1976). "Wright on functions." *Philosophical Review* 85:70-86.
- (1977). "Health as a theoretical concept." *Philosophy of Science* 44:542-73.
- (1987). "Concepts of health." In Donald VanDeVeer and Tom Regan, ed., *Health Care Ethics: An Introduction*. Philadelphia: Temple University Press.
- (1997). "A rebuttal on health." In J.M. Humber and Robert F. Almeder, ed., *What is Disease?* (Totowa, NJ: Humana Press), pp. 1-134.
- (2010). "Disability and medical theory." In D. Christopher Ralston and Justin Ho, eds, *Philosophical Reflections on Disability* (Dordrecht: Springer), pp. 55-88.
- (2011). "Concepts of health and disease." In Fred Gifford, ed., *Philosophy of Medicine*, vol. 12 of Elsevier *Handbook of Philosophy of Science*, (Oxford: Elsevier), pp. 13-64.
- (2014). "A second rebuttal on health." *Journal of Medicine and Philosophy* 39: 683-724.
- (2021). "Reznek on health." *Teorema* 40(1):23-65.
- (Under review). "Wakefield's critique of the biostatistical theory."
- Cooper, Rachel (2005). *Classifying Madness*. Dordrecht: Springer.
- (2015). "How might I have been?" *Metaphilosophy* 46:495-514.
- Crisp, Roger (2006). *Reasons and the Good*. Oxford: Clarendon Press.
- DeBlock, Andreas, and Sholl, Jonathan (2021). "Harmless dysfunctions and the problem of normal variation." In Faucher and Forest, ed., pp. 495-510.
- Dussault, Antoine C. (2018). "Welfare, health, and the moral considerability of nonsentient

biological entities.” *Ateliers de l'éthique/The Ethics Forum* 13:184-209.

----- (2020). “Wakefield’s harmful dysfunction account of disorder and the problem of defining harms to non-sentient organisms.” Unpublished manuscript.

----- (202x). “The harmful-dysfunction account of disorder, individual versus social values, and the interpersonal variability of harm challenge.” *Medicine, Health Care and Philosophy*, forthcoming.

Faucher, Luc, and Forest, Denis, ed. (2021). *Defining Mental Disorder: Jerome Wakefield and his Critics*. Cambridge, MA: The MIT Press.

Feinberg, Joel (1988). “Wrongful life and the counterfactual element in harming.” *Social Philosophy and Policy* 4:145-178.

Feit, Neil (2017). “Harm and the concept of mental disorder.” *Theoretical Medicine and Bioethics* 38:367-85.

FitzPatrick, William Joseph (2000). *Teleology and the Norms of Nature*. NY: Garland Publishing.

Fletcher, Guy (2016). *The Philosophy of Well-Being: An Introduction*. London: Routledge.

Foot, Philippa (2001). *Natural Goodness*. Oxford: Oxford University Press.

Fulford, K.W.M. (1989). *Moral Theory and Medical Practice*. NY: Cambridge University Press.

Geach, P.T. (1956). “Good and evil.” *Analysis* 17:23-42.

Horwitz, Allan V., and Wakefield, Jerome C. (2012). *All We Have to Fear*. NY: Oxford.

Hursthouse, Rosalind (1999). *On Virtue Ethics*. Oxford: Oxford University Press.

Kass, Leon (1975). “Regarding the end of medicine and the pursuit of health.” *The Public Interest* 40:11-42.

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

Lutz, Matthew. “Moral naturalism.” *Stanford Encyclopedia of Philosophy*.
www.plato.stanford.edu.

- McShane, Katie (2019). "Against etiological function accounts of interests." *Synthese*, published online June 27, 2019.
- Muckler, Dane, and Taylor, James Stacey (2020). "The irrelevance of harm for a theory of disease." *Journal of Medicine and Philosophy* 45:332-49.
- Nordenfelt, Lennart (1987). *On the Nature of Health*. Dordrecht: Reidel.
- (2007). "Holistic theories of health as applicable to non-human living beings." In H. Kincaid and J. McKittrick, ed., *Establishing Medical Reality* (NY: Springer).
- Nussbaum, Martha C. (2006). *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, MA: Belknap Press of Harvard University.
- Odenbaugh, Jay (2017). "Nothing in ethics makes sense except in the light of evolution? Natural goodness, normativity, and naturalism." *Synthese* 194:1031-1055.
- Parfit, Derek (1984). *Reasons and Persons*. Oxford: Clarendon Press.
- Powell, Russell, and Scarffe, Eric (2019a). "Rethinking 'disease': a fresh diagnosis and a new philosophical treatment." *Journal of Medical Ethics* 45(9):579-88.
- (2019b) "Rehabilitating 'disease': function, value, and objectivity in medicine." *Philosophy of Science* 86:1168-78.
- Rachels, James (2019). *The Elements of Moral Philosophy*, 9th ed. NY: McGraw-Hill.
- Reznek, Lawrie (1991). *The Philosophical Defense of Psychiatry*. London: Routledge and Kegan Paul.
- Rice, Christopher M. (2015). "Well-being and animals." In Guy Fletcher, ed., *The Routledge Handbook of Philosophy of Well-being* (Routledge), pp. 378-88.
- Sedgwick, Peter (1973). "Illness – mental and otherwise." *Hastings Center Studies* 1:19-40.
- Singer, Peter (1994). *Rethinking Life and Death*. NY: St. Martin's Griffin.
- Sommerhoff, Gerd (1950). *Analytical Biology*. London: Oxford University Press.
- (1959). "The abstract characteristics of living organisms." In F.E. Emery, ed., *Systems Thinking* (Harmondsworth: Penguin), pp. 147-202.

Spitzer, R.L. and Endicott, J. (1978). "Medical and mental disorder: proposed definition and criteria." In Spitzer and R.L. Klein, ed., *Critical Issues in Psychiatric Diagnosis* (NY: Raven Press).

Spitzer, R.L. and Williams, J.B.W. (1982). "The definition and diagnosis of mental disorder." In W.R. Gove, ed., *Deviance and Mental Illness* (Beverly Hills, CA: Sage), pp. 15-31.

Sumner, L.W. (1996). *Welfare, Happiness, and Ethics*. NY: Oxford University Press.

Taylor, Paul W. (1986). *Respect for Nature: A Theory of Environmental Ethics*. Princeton: Princeton University Press.

Thomson, Judith Jarvis (2001). *Goodness and Advice*. Princeton, NJ: Princeton University Press

Thompson, Michael (1995). "The representation of life." In R. Hursthouse, G. Lawrence, and W. Quinn, ed., *Virtues and Reasons* (Oxford: Clarendon Press).

Varner, Gary E. (1998). *In Nature's Interests? Interests, Animal Rights, and Environmental Ethics*. NY: Oxford University Press.

----- (2003). Review of Nicholas Agar, *Life's Intrinsic Value: Science, Ethics, and Nature*. *Environmental Ethics* 25:413-16.

von Wright, G. H. (1963). *The Varieties of Goodness*. London: Routledge and Kegan Paul.

Wakefield, Jerome C. (1992a). "The concept of mental disorder: on the boundary between biological facts and social values." *American Psychologist* 47:373-88.

----- (1992b). "Disorder as harmful dysfunction: a conceptual critique of DSM-III-R's definition of mental disorder." *Psychological Review* 99:232-47.

----- (1997). "Diagnosing DSM-IV, part 1: DSM-IV and the concept of mental disorder." *Behaviour Research and Therapy* 35:633-49.

----- (1999). "Mental disorder as a black box essentialist concept." *Journal of Abnormal Psychology* 108:465-72.

----- (2000). "Spandrels, vestigial organs, and such: reply to Murphy and Woolfolk's 'The harmful dysfunction analysis of mental disorder'." *Philosophy, Psychiatry, and Psychology* 7:253-69.

----- (2006). "Personality disorder as harmful dysfunction: DSM's cultural deviance criterion reconsidered." *Journal of Personality Disorders* 20(2):157-69.

----- (2013). "Addiction, the concept of disorder, and pathways to harm: comment on Levy." *Frontiers in Psychiatry* 4:1-2.

----- (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* 39:648-82.

----- (2017). "Addiction and the concept of disorder, part I: why addiction is a medical disorder." *Neuroethics* 10:39-53.

----- (2021a). "From Ribot and Dupre to Spitzer and RDoC: Does the harmful dysfunction analysis possess historical explanatory power? Reply to Steeves Demazeux." In Faucher and Forest, ed., pp. 27-46.

----- (2021b). "Is the harmful dysfunction analysis descriptive or stipulative, and is the HDA or BST the better naturalist account of dysfunction: reply to Mael Lemoine." In Faucher and Forest, ed., pp. 213-50.

----- (2021c). "Does developmental plasticity pose a challenge to the harmful dysfunction analysis? Reply to Justin Garson." In Faucher and Forest, ed., pp. 353-84.

----- (2021d). "Biological function hierarchies and indeterminacy of dysfunction: supplementary reply to Justin Garson." In Faucher and Forest, ed., pp. 385-96.

----- (2021e). "Do the challenges of autism and neurodiversity pose an objection to the harmful dysfunction analysis? Reply to Denis Forest." In Faucher and Forest, ed., pp. 449-68.

----- (2021f). "Can the harmful dysfunction analysis distinguish problematic normal variation from disorder? Reply to Andreas De Block and Jonathan Sholl." In Faucher and Forest, ed., (2021), pp. 511-36.

----- (2021g). "Must social values play a role in the harm component of the harmful dysfunction analysis? Reply to Rachel Cooper." In Faucher and Forest, ed. (2021), pp. 553-75.

----- and Conrad, Jordan A. (2019). "Does the harm component of the harmful dysfunction analysis need rethinking? Reply to Powell and Scarffe." *Journal of Medical Ethics* 45(9):594-6.

----- (2020). "Harm as a necessary component of the concept of medical disorder: reply to Muckler and Taylor." *Journal of Medicine and Philosophy* 45:350-70.

Wakefield, Jerome C., and Horwitz, Allan V. (2007). *The Loss of Sadness*. NY: Oxford.

World Health Organization (vol 1 1992, vol 2 1993, vol 3 1994). *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision [ICD-10]*. Geneva: WHO.