# A Clarification on the Boorse-Wakefield Debate about Health: Is the Theoretical/Therapeutic Distinction Dispensable?

**Antoine C. Dussault**

**Affiliation 1:** Collège Lionel-Groulx, Sainte-Thérèse, Québec, Canada

**Affiliation 2:** Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Montréal, Québec, Canada

**Email:** Antoine.CDussault@clg.qc.ca

ORCID: 0000-0003-2164-1571

## Abstract

Although Boorse's and Wakefield's accounts of health are generally regarded as competing ones, they are in fact so only if they are aimed at the same concept. Some remarks made by Boorse and Wakefield, however, leave it unclear whether they are. On one possible interpretation, Boorse's account aims at analyzing a *theoretical* concept of abnormality, which ought to be distinguished from a more *clinical* or *therapeutic* concept, whereas Wakefield's account aims at analyzing a *clinical* or *therapeutic* concept. The debate between Boorse and Wakefield would then either be merely terminological, or would boil down to whether Boorse is correct to assert the existence of a *theoretical* concept of abnormality which ought to be distinguished from a *clinical* or *therapeutic* one. This paper aims to clarify what is at stake between Boorse and Wakefield, by maintaining that their accounts are most plausibly interpreted as both being aimed towards a *theoretical* concept of abnormality.

## 1. Introduction

An important matter of contention between Christopher Boorse's (1977; 1997; 2014) and Jerome Wakefield's (1992b; 2014) accounts of health concerns whether part-dysfunction should or should not be considered as sufficient for decreasing an individual's health.[1] Whereas Boorse calls such decreases of health *pathologies* and equates them with part-dysfunctions, Wakefield calls such decreases *medical disorders* and restricts the class of medical disorders to *harmful* dysfunctions, that is, dysfunctions that cause harm to their carrier.

Although Boorse's and Wakefield's accounts are generally regarded (even by themselves) as competing ones, they are in fact so only if they purport to provide definitions of the same concept. Some remarks made by Boorse and Wakefield, however, leave it unclear whether they do. That is, they leave it unclear whether *pathology*, the target of Boorse's analysis, is equivalent to *disorder*, the target of Wakefield's analysis. The remarks at issue have to do with the distinction Boorse introduces between *theoretical* and *clinical* or *therapeutic* abnormality, which he argues that Wakefield erroneously ignores (Boorse 1997: 48–49), and of which Wakefield is at times skeptical (Wakefield 2014: 660).

[1] Another difference between Boorse's and Wakefield's accounts concerns the understanding of function on which each account is based. Boorse adopts a *goal-contribution* account of function (e.g. Boorse 1976), whereas Wakefield adopts a version of the *selected effects* account of function (e.g. Neander 1991; Wakefield 2005). Yet another difference is that Boorse is committed to analyzing the concept of pathology used by medical professionals only, while Wakefield seeks to account for both professional and lay people's concepts (see Wakefield 2014: 652). These two further differences need not concern us here.

In light of Boorse's theoretical/therapeutic distinction, the difference between Boorse and Wakefield's accounts may look very thin: Boorse defines pathology as part-dysfunction, but recognizes the existence of *therapeutic* abnormality, that is, pathology that requires treatment, and this latter concept seems very close to Wakefield's concept of harmful dysfunction. If this were so, then the debate between Boorse and Wakefield would either be merely terminological, or would boil down to whether Boorse is correct in asserting the existence (or medical importance) of a *theoretical* concept of abnormality that ought to be distinguished from a *therapeutic* concept. Boorse could then dismiss Wakefield's criticism of his account by arguing that it rests on a confusion between the theoretical and therapeutic concepts of abnormality, and Wakefield could strike back by casting doubts on Boorse's theoretical/therapeutic distinction.[2]

[2] The reading that sees no substantial disagreement between Boorse and Wakefield may also seem to be reinforced by the observation that Wakefield (2014: 654–55) readily recognizes the legitimacy of a concept of pathology defined as part-dysfunction—provided, however, that "pathology" is applied *strictly to the dysfunctional parts* themselves and not to the individual who carries them. This observation suggests that, in a sense, Boorse and Wakefield both recognize the validity of the concept targeted by the other's theory: Wakefield's "part pathology" concept equates to Boorse's strictly naturalistic part-dysfunction concept of pathology, and conversely, as suggested above, Boorse's therapeutic abnormality equates to Wakefield's concept of harmful dysfunction. The equivalence between Wakefield's part pathology concept and Boorse's pathology-as-part-dysfunction concept, however, is not valid, for two reasons. First, for Boorse, a part-dysfunction constitutes not just a pathology *of the dysfunctional part* (what Wakefield acknowledges), but also a pathology *of the individual* who carries the part. Second, for Boorse, the pathology-as-part-dysfunction concept is *the* central medical concept, whereas the part pathology concept plays no important medical role for Wakefield. These are two substantial disagreements that remain despite Wakefield's

This paper aims to clarify what is at stake between Boorse and Wakefield, by determining whether Wakefield's concept of harmful dysfunction should or should not be regarded as equivalent to Boorse's concept of therapeutic abnormality. I will argue that the two concepts *cannot* be equated, because Wakefield cannot dispense with a concept of therapeutic abnormality as clearly distinguished from a theoretical concept. Wakefield's account must then, like Boorse's, be understood as aimed towards a *theoretical* concept of abnormality. This implies that Wakefield cannot dismiss Boorse's account simply by casting doubt upon the theoretical/therapeutic distinction, but, on the other hand, that Boorse cannot dismiss Wakefield's criticism simply by arguing that it conflates theoretical and therapeutic abnormality. The debate between Boorse and Wakefield must then be about which of their accounts provides the most adequate definition of *theoretical abnormality*.

## 2. Benign dysfunctions and the theoretical/therapeutic distinction

The issue of whether Boorse and Wakefield's analyses are aimed at the same target concept arises in the context of their discussion of a type of purported counterexample to Boorse's account: benign dysfunctions. According to Wakefield (2014), benign dysfunctions raise a challenge for Boorse's account, because they constitute cases that this account seems wrong to classify as pathologies. Insofar as, by definition, benign

---

acknowledgement of a part pathology concept. The remainder of this paper will focus on the other suggested equivalence: that between Boorse's therapeutic or clinical abnormality and Wakefield's harmful dysfunction concept. All along, it should be understood that, when I speak of Wakefield as rejecting part-dysfunction's sufficiency for pathology, what I mean is part-dysfunction's sufficiency for pathology *of the individual*.

dysfunctions cause no further problems to the individuals who carry them, Wakefield argues that these dysfunctions should not be considered pathological (or disordered). Wakefield mentions as examples of such benign dysfunctions the lack of one kidney after transplantation, the asymptomatic carriage of infectious agents, asymptomatic HIV infection, the carriage of neutral, risky or benign mutations, *situs inversus totalis*, benign angiomas, etc. (Wakefield 2014: sec. 4; Wakefield and Conrad 2020: 357–59; Wakefield 2021a: 518–20).

In response to a similar challenge previously raised by Lennart Nordenfelt (1987), Boorse (1997) invokes his distinction between theoretical and therapeutic abnormality. Nordenfelt's challenge focuses on the case of a person carrying just *one dead cell*, or one cell that is unable to perform its normal function. Nordenfelt (1987: 28) points out that Boorse's theory must, wrongly in his view, consider the dead or malfunctioning cell as a pathological condition. In response, Boorse bites the bullet, and claims that he can disarm Nordenfelt's challenge by attributing it to a confusion between theoretical and therapeutic abnormality:

> At the pathologist's level of description, there is no paradox in calling one dead cell pathological, except, of course, in tissues like skin and mucosa whose normal function entails constant death and regeneration. One dead cell is just the ultimate in focal necrosis, one of pathology's most common findings. … Is a dead neuron a normal neuron? Do dead neurons function normally? Since the neuron is a body part, in a tissue whose physiological function does not feature regular regeneration, it is not much more mysterious that its death is pathological than that the whole organism's death is pathological. … Of course, one dead neuron is a trivial piece of pathology. But to call a condition pathological implies nothing about its importance. To think otherwise is to confuse *theoretical* and *clinical* normality (Boorse 1997: 50–51, italics added).

Boorse here speaks of *clinical* (ab)normality, but the context makes it clear that he refers to the subset of clinical abnormality he calls *therapeutic* abnormality.

Boorse first introduced his theoretical/therapeutic distinction in a previous paper (Boorse 1987: 365), where he contrasts theoretical abnormality with other, more clinically relevant, concepts of abnormality:

- *Diagnostic abnormality*: Conditions that are theoretically abnormal (i.e. pathological), and clinically detectable. This excludes clinically undetectable pathological conditions, such as tiny pancreatic cysts and transient cardiac arrhythmias.

- *Therapeutic abnormality:* Conditions that are theoretically and diagnostically abnormal (i.e. pathological and clinically detectable), and that necessitate treatment. This excludes clinically detectable pathological conditions that do not necessitate treatment, such as benign tumors and small skin lesions.

- *Death*: The interruption of all biological functions (i.e. the most extreme case of pathology).

"Importance" in Boorse's response to Nordenfelt seems to refer to whether a pathological condition is severe enough to warrant treatment. If this is so, then the clinical concept of abnormality he has in mind when making this response must be his *therapeutic* abnormality. In his 1997 paper, Boorse labels the above clinical concepts of abnormality "disease-plus" concepts (Boorse 1997: 55, 100).

Wakefield (2014) criticizes Boorse's appeal to the theoretical/therapeutic distinction in response to the one-dead-cell objection as ad hoc, because Boorse provides no independent reasons to accept it. Wakefield even goes so far as to assert that Boorse's "claimed distinction between the pathologist's and the clinician's concepts of disorder does not exist." (Wakefield 2014: 660) If this were the case, then it would seem more natural to accept Wakefield's harmful dysfunction analysis, which excludes from the outset the one dead cell and other benign dysfunctions, than to adopt Boorse's definition and then be forced to multiply abnormality or "disease-plus" concepts to deal with such dysfunctions.

This could be the end of the debate if Wakefield could really dispense with the theoretical/therapeutic distinction introduced by Boorse (and if Boorse proved unable to provide further justification for this distinction). However, Wakefield can dispense with the theoretical/therapeutic distinction only if what he means by "harmful dysfunction" can be equated with Boorse's notion of therapeutic abnormality, that is, if the target concept of his harmful dysfunction analysis is equivalent to Boorse's therapeutic concept (i.e. dysfunctions that warrant treatment).

Whether this is so, however, seems to be an unsettled matter between him and Boorse. As I just mentioned, Wakefield asserts that Boorse's "claimed distinction between the pathologist's and the clinician's concepts of disorder does not exist" (Wakefield 2014: 660). This suggests that he envisions his harmful dysfunction definition as aimed towards a therapeutic concept that does not need to be complemented by a theoretical concept. Such a reading seems to be confirmed by Wakefield's suggestion that Boorse's clinical concept corresponds "perhaps to 'harmful dysfunction'" (Wakefield

2014: 651). Elsewhere, however, Wakefield recognizes that "the status of a condition as disordered or nondisordered from the HD [i.e. his harmful dysfunction account] or any other perspective has no necessary implication for the priority the condition deserves with respect to treatment, prevention, or policy" (Wakefield 1999: 374). He also asserts that "a correct definition of disorder must classify every pathological condition as a disorder whether or not the condition is currently an object of professional attention" (Wakefield 1992a: 234). The latter statements seem to align with the spirit of Boorse's theoretical/therapeutic distinction, by admitting that not all disorders require treatment, and that this is so even when we define disorder as harmful dysfunction.

Along similar lines, Boorse, in his 1997 paper, casts Wakefield's harmful dysfunction account as aimed towards his therapeutic concept of abnormality. He states:

> Wakefield's paper illustrates that this point and the pathological/clinical contrast are no small matters. Wakefield sets out to define "disorder" for purposes of psychiatric classifications. But like Spitzer and other participants in the DSM-III project, Wakefield means disorder to be a clinical concept. … Accordingly, both Wakefield and Spitzer ignore the distinction between pathological and clinical concepts, and that is why Wakefield feels he must supplement dysfunction with a harm clause … What he is analyzing is really "clinical disease" (more exactly, "therapeutic abnormality"), with dysfunction analyzing the "disease" part and harm analyzing the "clinical" part (Boorse 1997: 48–49).

However, in a later paper, Boorse takes note of the above-quoted passage from Wakefield (1999), and reads it (as I do) as indicating that Wakefield "agrees with the BST that judgments of pathology entail no therapeutic or social ones" (Boorse 2011: 35).[3]

In the next section, I will argue that certain aspects of Wakefield's construal of harm require him to recognize Boorse's theoretical/therapeutic distinction.

## 3. Harmful dysfunctions vs. treatment-requiring dysfunctions

Can Wakefield dispense with the theoretical/clinical distinction introduced by Boorse? We have seen that, as Boorse construes it, the main purpose of this distinction is to leave open the possibility that not all pathological conditions be considered serious enough to require treatment. Wakefield then seems able to dispense with the theoretical/therapeutic distinction only if he can maintain that *all* disorders as defined by his account necessitate treatment. That is, he must be able to maintain that dysfunctions that are harmful in the sense relevant to his account must always be treated (in contradiction with what he sometimes seems to assert; see section 2 above). Otherwise, Wakefield will find himself implicitly acknowledging his own version of the theoretical/therapeutic distinction, according to which *theoretical* abnormalities will be harmful dysfunctions, and *therapeutic* abnormalities will be the subset of these harmful dysfunctions that necessitate treatment.

[3] Though Boorse cautiously notes that "this may just mean that [for Wakefield] disorders, though always prima facie worthy of treatment, need not be so on balance."

Wakefield's construal of the notion of harm that he considers relevant to his theory forces him, it seems, to recognize at least two categories of disorders that do not always necessitate treatment. The first category results from his adoption of a social-values-based construal of harm. As Wakefield states in the paper in which he first introduced his harmful dysfunction account, "a disorder exists when the failure of a person's internal mechanisms to perform their functions as designed by nature impinges harmfully on the person's well-being *as defined by social values and meanings*" (Wakefield 1992b: 373, italics added). According to Wakefield, then, a condition is harmful in the sense relevant to his account when the cultural value system of the group to which an individual belongs implies that it is harmful, irrespective of whether the affected individual herself considers it harmful. As he illustrates in more recent publications, this implies that a North American person with a dysfunction causing infertility is in a disordered state even if she sees her infertility as advantageous because she does not desire to conceive children (Wakefield and Conrad 2019: 1; Wakefield 2021b: 557). This is because North American culture values the ability to conceive children and hence perceives the inability to do so as harmful.

So conceiving of harms that qualify a dysfunction as a disorder seems to force Wakefield to recognize a first category of disorders (i.e., harmful dysfunctions) that need not require treatment: disorders that are acceptable to the affected individual despite being regarded as harmful within her sociocultural group. For instance, if a North American person sees her infertility as advantageous because she does not desire to conceive children, then it would seem reasonable for her to prefer not to treat her condition (provided it causes her no further harm), irrespective of whether her culture

considers it harmful (Wakefield explicitly recognizes this, see Wakefield 2021a: 513, 515; 2021b: 563).

A second category of disorders that need not require treatment arises from a clarification Wakefield recently made that harms that qualify a dysfunction as a disorder in the sense relevant to his account are *pro tanto harms* rather than *all things considered* harms (Wakefield and Conrad 2019: 1–2; 2020: sec. III). What he means by this is that the harms that qualify a dysfunction as a disorder include those that are offset by indirect benefits. For example, a broken arm remains harmful and therefore a disorder according to Wakefield's account even when the annoyance of having one's arm broken is offset by the benefit of being able to take time off work or receive insurance payments. Likewise, cowpox remains harmful and therefore a disorder according to Wakefield's account even during a smallpox epidemic, when the harm it causes is offset by its provision of an advantageous resistance to smallpox.

This inclusion of pro tanto harms that are offset by indirect benefits among those that qualify a dysfunction as a disorder seems to force Wakefield to admit a second category of disorders (i.e., harmful dysfunctions) that need not require treatment: disorders that provide a net benefit that would be lost if the affected individual were treated. For instance, since cowpox provides an advantageous resistance to smallpox, then it would seem reasonable for a patient affected with cowpox during a smallpox epidemic to prefer not to treat her condition.

It seems, therefore, that Wakefield cannot dispense with Boorse's contrast between theoretical and therapeutic abnormalities. Even if he restricts the class of disorders to

harmful dysfunctions, he nevertheless finds himself constrained to distinguish, among disorders, those that necessitate treatment and are therefore *therapeutically abnormal* from those that do not necessitate treatment and are therefore therapeutically normal, although *theoretically abnormal*. Wakefield, indeed, could reject the theoretical vs. therapeutic *terminology*, but he seems compelled to accept the contrast this terminology is intended to draw.

Wakefield could, of course, escape this conclusion if he changed his construal of harm. Doing so, however, might have costly implications. For instance, giving up his social-values-based view of harm, and instead adopting a view of harm as defined by patients' individual assessments of their condition, would force him to accept that whether a condition is a disorder may vary across individuals, insofar as assessments of a given condition's harmfulness typically varies across individuals. This seems to be an implication that he is not willing to accept (see Wakefield 2021a: 513). In the current state of his account, at least, Wakefield seems compelled to accept the spirit of Boorse's theoretical/therapeutic distinction, that is, the view that not all pathologies or disorders—i.e. *theoretical* abnormalities—always necessitate treatment—i.e. are also *therapeutic* abnormalities.

## 4. Conclusion

I argued above that Wakefield cannot dispense with a concept of therapeutic abnormality as clearly distinguished from a theoretical concept. This entails that the debate between him and Boorse must be understood as one about which of the two philosophers' accounts provides the most adequate definition of *theoretical abnormality*.

Wakefield cannot dismiss Boorse's account simply by casting doubt upon the theoretical/therapeutic distinction. Since he is implicitly committed to this distinction, he must argue that a definition of theoretical abnormality along the lines of his account is more plausible than one along the lines of Boorse's account. Likewise, Boorse cannot dismiss Wakefield's criticism simply by arguing that it conflates theoretical and therapeutic abnormality. He must argue that, once we have recognized the theoretical/therapeutic distinction, we should adopt a definition of theoretical abnormality along the lines of his own account rather than one along the lines of Wakefield's account.

The issue thus turns out to be over which of the two following ways of slicing up medical concepts is the most adequate one:

*A Boorsean slicing-up:* A concept of theoretical abnormality defined in purely functional-biological terms, *plus* an array of clinical or "disease-plus" concepts—diagnostic abnormality, therapeutic abnormality, death—to which a concept of *harmful abnormality* (or *harmful disease*, as Boorse 1997: 100, suggests), equivalent to harmful dysfunction, might be added; or

*A Wakefieldian slicing-up:* A concept of theoretical abnormality defined as harmful dysfunction, *plus* an array of clinical or "disease-plus" concepts (that at least includes therapeutic abnormality), with no recognition of any important medical role for a purely biological concept of part-dysfunction.[4]

[4] In line with what is said in footnote 2, part-dysfunction would at best define a (medically insignificant) concept of part pathology.

I leave it open here which of these two slicings-up, or a possible third one, offers the most adequate picture.[5]

## Funding

The work for this paper was supported by a research grant from the Fonds de recherche du Québec – Société et culture (FRQSC, 2018-CH-211053).

## References

Boorse, Christopher. 1976. "Wright on Functions." *Philosophical Review* 85 (1): 70–86.
———. 1977. "Health as a Theoretical Concept." *Philosophy of Science* 44 (4): 542–73.
———. 1987. "Concepts of Health." In *Health Care Ethics: An Introduction*, edited by Donald VanDeVeer and Tom Regan, 359–93. Philadelphia: Temple University Press.
———. 1997. "A Rebuttal on Health." In *What Is Disease?*, edited by James M. Humber and Robert F. Almeder, 1–134. Totowa, N.J.: Humana Press.
———. 2011. "Concepts of Health and Disease." In *Philosophy of Medicine*, edited by Fred Gifford, 13–64. Amsterdam: Elsevier.
———. 2014. "A Second Rebuttal On Health." *Journal of Medicine and Philosophy* 39 (6): 683–724.
Neander, Karen. 1991. "Functions as Selected Effects: The Conceptual Analyst's Defense." *Philosophy of Science* 58 (2): 168–84.
Nordenfelt, Lennart Y. 1987. *On the Nature of Health: An Action-Theoretic Approach*. Dordrecht, the Netherlands: D. Reidel.
Wakefield, Jerome C. 1992a. "Disorder as Harmful Dysfunction: A Conceptual Critique of DSM-III-R's Definition of Mental Disorder." *Psychological Review* 99 (2): 232–47.
———. 1992b. "The Concept of Mental Disorder: On the Boundary between Biological Facts and Social Values." *American Psychologist* 47 (3): 373–88.
———. 1999. "Evolutionary versus Prototype Analyses of the Concept of Disorder." *Journal of Abnormal Psychology* 108 (3): 374–99.

[5] I am thankful to Christopher Boorse and two anonymous referees for their valuable comments and suggestions, and to Luc Faucher for asking me the question that triggered my thinking on the issue discussed in this paper. I also thank Alice Everly for editing my manuscript.

———. 2005. “Biological Function and Dysfunction.” In *Handbook of Evolutionary Psychology*, edited by David M. Buss, 878–902. Hoboken, N. J.: Oxford University Press.

———. 2014. “The Biostatistical Theory Versus the Harmful Dysfunction Analysis, Part 1: Is Part-Dysfunction a Sufficient Condition for Medical Disorder?” *Journal of Medicine and Philosophy* 39 (6): 648–82.

———. 2021a. “Can the Harmful Dysfunction Analysis Distinguish Problematic Normal Variation from Disorder? Reply to Andreas De Block and Jonathan Sholl.” In *Defining Mental Disorder: Jerome Wakefield and His Critics*, edited by Luc Faucher and Denis Forest, 511–36. Cambridge, Mass.: MIT Press.

———. 2021b. “Must Social Values Play a Role in the Harm Component of the Harmful Dysfunction Analysis? Reply to Rachel Cooper.” In *Defining Mental Disorder: Jerome Wakefield and His Critics*, edited by Luc Faucher and Denis Forest, 553–75. Cambridge, Mass.: MIT Press.

Wakefield, Jerome C., and Jordan A. Conrad. 2019. “Does the Harm Component of the Harmful Dysfunction Analysis Need Rethinking?: Reply to Powell and Scarffe.” *Journal of Medical Ethics* 45 (9): 594–96. https://doi.org/10.1136/medethics-2019-105578.

———. 2020. “Harm as a Necessary Component of the Concept of Medical Disorder: Reply to Muckler and Taylor.” *Journal of Medicine and Philosophy* 45 (3): 350–70.