

Philosophy of Medicine

Book Review

Medical Disorder Is Not a Black Box Essentialist Concept Review of *Defining Mental Disorder: Jerome Wakefield and His Critics*, edited by Luc Faucher and Denis Forest

Harriet Fagerberg¹

¹ Hunter College and the Graduate Center, City University of New York, NY, USA. Email: harriet.fagerberg@hunter.cuny.edu

1. Introduction

Defining Mental Disorder: Jerome Wakefield and His Critics, edited by Denis Forest and Luc Faucher, is essential reading for students and researchers in philosophy of medicine whose work is informed by that of Jerome Wakefield, or the disease debate in general. If you are anything like me, this book will open the door to a new depth of understanding of the harmful dysfunction analysis (HDA) and its methodical underpinnings, and an enriched appreciation of what is at stake in defining medical disorder. To those confident that they understand Wakefield's view and the surrounding debate, thinking perhaps that they do not need a refresher, I would say: perhaps you do not understand it as well as you think. *Defining Mental Disorder* provides much more than an overview of previously well-trodden ground. The chapters themselves, and Wakefield's own detailed responses, help to elucidate a number of methodological nuances and peculiarities that have rarely been covered in such depth before. I was startled to learn how philosophically complex Wakefield's position actually is, and how shallow my own understanding of it has been. It emerges as a far more nuanced and sophisticated account than I had given it credit for, and one that deserves serious reevaluation.

2. Summary

In Wakefield's influential conceptual analysis of medical disorder, disorders are just harmful dysfunctions—*bad* failures to function as we were *designed* to function by evolution. To me, the intuitive appeal of Wakefield's view has always been obvious. The harm criterion has commonsense appeal, and the evolutionary theory of biological function has a strong theoretical and scientific rationale. Contrast this with Christopher Boorse's biostatistical theory, which leans on complicated statistical criteria, paired with an idiosyncratic theory of proper function, which has had little uptake in the function literature (see Boorse 1977; cf. Garson 2019). However, in the course of making my way through *Defining Mental Disorder*, I have come to appreciate that Wakefield's HDA is idiosyncratic



This work is published by [Pitt Open Library Publishing](#) and is licensed under a [Creative Commons Attribution 4.0 International License](#). © The Author(s).

in its own, more subtle way and, indeed, widely misunderstood. Wakefield is admirably clear about his methodological objectives throughout his contributions to this volume, more so perhaps than ever before. As he writes in his introduction: “For those readers familiar with the HDA, I should mention that the critics’ compelling arguments have moved me to alter or amplify or clarify my view on several issues. The HDA survives intact but in a more nuanced and elaborated form” (Faucher and Forest 2021, xxvii). In particular, while his proposed criteria survive unchanged, Wakefield does much in this volume to clarify his methodological approach. From his incredibly thorough and in-depth responses to his critics, it becomes clear that many critiques of the HDA are misguided and rest on misinterpretations of Wakefield’s view. To my mind, he has succeeded in demonstrating this to the reader. However, it also emerges that Wakefield’s view is very modest and plausible in one respect, which arguably has received disproportionate philosophical scrutiny, while being extremely philosophically demanding in ways that have hardly been debated at all. In what follows, I first summarize the volume as a whole, before outlining Wakefield’s view as I now understand it.

The volume is split into four parts. As the editors explain, each part deals with a distinct aspect of Wakefield’s HDA. Part I concerns his chosen method of inquiry: conceptual analysis. It contains four chapters, each followed by a chapter-length response from Wakefield. First, Steeves Demazeux considers the historical significance of the “dysfunction” component of the HDA, questioning its centrality through time. Then Luc Faucher challenges Wakefield to put his claims to empirical test, using the methods of experimental philosophy. The exchange between Wakefield and Faucher is followed by Leen De Vreese making the case for a pluralism about medical concepts. Finally, Harold Kincaid questions whether we really need a concept of “mental disorder” in the study of psychopathology.

Part II, which contains only a single chapter contributed by a critic, again followed by Wakefield’s response, concerns the demarcation problem in philosophy of medicine. Peter Zachar questions Wakefield’s “essentialism” about mental disorder in the context of demarcation.

Part III considers the dysfunction component of Wakefield’s HDA, and contains the bulk of the volume’s content, with six chapters contributed by critics on this topic, each with a response from Wakefield, and two receiving a supplementary response as well. First, Maël Lemoine considers a tension in Wakefield’s goal of providing a descriptively adequate account of the meaning of mental disorder, and one that serves our scientific purposes. Dominic Murphy then questions Wakefield’s commitment to an evolutionary theory of function. Murphy is followed by Justin Garson, who raises a novel challenge to the dysfunction criterion from the possibility of “developmental plasticity.” Philip Gerrans argues that instead of analyzing disorders in terms of dysfunction of “mental mechanisms,” we should look more closely at the functions of neural and molecular mechanisms. Denis Forest goes on to challenge Wakefield’s account, with autism as a case study, followed by Tim Thornton comparing Wakefield’s project with analogous projects in the philosophy of mental content.

Finally, Part IV considers the harm component of Wakefield’s analysis. Andreas De Block and Jonathan Sholl argue that the distinction between harm and dysfunction is not as neat as Wakefield claims. Rachel Cooper then challenges Wakefield on his conception of “harm” as whatever is disvalued by society. Again, each author gets a chapter-length

response from Wakefield, with Cooper receiving an additional supplementary response on the possibility of selected disorders.

3. Wakefield's Harmful Dysfunction Analysis

Wakefield has always been clear that his method is the “conceptual analysis” of the “concept” of medical and in particular mental disorder (see, for example, Wakefield 2014). The focus on method in this volume is nonetheless welcome because both “concepts” and “conceptual analysis” are, to some degree, open to interpretation. Some think of concepts as internal psychological states that guide our practices of classification, while some think concepts are externally individuated (see, for example, Machery 2009; cf. Millikan 2000). Some think of conceptual analysis as, simply, the practice of describing widely shared psychological states that guide our language use, while some think conceptual analysis has *always* referred to substantive attempts to understand real kinds, properties, and phenomena in the external world (see Neander 1991; cf. Deutsch 2021; see also Papineau 2009).

The HDA, as originally proposed by Wakefield, states that a condition is a disorder if it “causes some harm or deprivation of benefit to the person as judged by the standards of the person’s culture” and “results from the inability of some internal mechanism to perform its natural function, wherein a natural function is an effect that is part of the evolutionary explanation of the existence and structure of the mechanism” (Wakefield 1992, 384). The criteria themselves have survived unchanged from his early work. As Wakefield consistently describes his view: “The HDA 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” (Faucher and Forest 2021, 27). However, throughout this volume, Wakefield elaborates on his view in important respects when it comes to the methodological framing of his proposed criteria.

On the one hand, Wakefield’s view is a conceptual analysis in the most traditional and straightforward sense. All he is doing is offering an empirically refutable hypothesis. If some competent, fully informed member of the relevant linguistic community—let us call them Jane—thought that there was something going *wrong* inside another individual (Joe), and judged moreover that this “going wrong” was also *bad* in accordance with the standards of the prevailing culture, then Jane would also judge that Joe had a medical disorder. And vice versa—if Jane found that what was going on inside Joe was *not* a case of something going awry organismically, or not harmful according to our social standards, then she would *not* judge that Joe had a medical disorder. As Wakefield puts it: “My claim—and the prediction that follows from the HDA—is that when a condition believed to be a disorder is found to [not be a dysfunction] there will be a tendency to reclassify the condition as a nondisorder” (Faucher and Forest 2021, 361).

So the HDA is a thesis, not about what is in the world but about what is in our heads. Wakefield’s implicit view of concepts is one in which conceptual content is determined by internal psychological states. In principle, then, Wakefield offers an empirically refutable psychological hypothesis about the belief-states that guide us in judging that some state or condition qualifies as an instance of pathology. As such, it is, in principle, falsifiable. If it turned out that we were inclined to use disease terms (disease, disorder, pathology, and so

on) to describe cases absent of something going biologically wrong in a socially bad way, or vice versa, then Wakefield's hypothesis would, *ipso facto*, be shown to be mistaken. In his chapter, Garson praises Wakefield's commitment to an admirably risky bet in this respect. However, there are three possible sources of wiggle room for Wakefield that render his view less exposed.

Firstly, it is clear that "meaning is in the head" for Wakefield but it is not quite clear whose head—lending some imprecision to Wakefield's empirical prediction. Whose judgments would we have to test? In previous work, he has been explicit that his account departs from that of Boorse (who claims to be interested, mainly, in the judgments of medical pathologists) in that he includes laypeople as well as clinicians in the community from which evidence is drawn (Wakefield 2014). However, in his response to Kincaid, he insists that "precisely specifying a preset target community is not necessary" because "the 'community' is in effect a construct of those dispersed throughout the professional and lay communities who share a certain widespread concept of disorder that is a salient one among the many in circulation and has certain properties that make it important in scientific research and lay debate" (Faucher and Forest 2021, 139).¹ Wakefield's position here is a bit mysterious to me and borders, perhaps, on the circular. "The community" are those people who possess the concept, in Wakefield's sense?²

Wakefield's strict distinction between the analytic and the synthetic provides a second source of possible leeway. He is very clear in his response to critics that his empirical bet is relative to the idiosyncratic background assumptions and beliefs of the language users in question. Two people might share a concept and yet be inclined to apply it rather differently because they have different understandings of the relevant empirical facts. As Wakefield writes in his response to De Vreese:

Concepts by themselves do not determine classificatory judgments. You and a friend can totally agree on the meaning of the term "bachelor" and totally disagree about whether a specific individual you met at a party is or is not a bachelor based on different observations, beliefs, and inferences ... The concept and the presuppositions are jointly necessary for explaining classificatory judgments. (Faucher and Forest 2021, 112)

This opens up another route for explaining away apparent applications of the term "disorder" that defy the HDA. When competent users of the concept apply terms like "disease" and "disorder" to conditions where it seems, to us, that there is no harm or no biological dysfunction, Wakefield can readily explain this in non-semantic terms. Surely, they must *believe* that there is harm and dysfunction. Why else would they apply "disorder" in this mistaken way?

Third, note that, at this stage, "failure to function as one should" is a vague notion. Although Wakefield is committed to an evolutionary or etiological theory of function—in which functions are effects for which a trait has been naturally selected—this *theory* of

¹ The HDA, Wakefield explains, can alternatively be thought of as a "transcendental argument," sensitive to certain normative desiderata (Faucher and Forest 2021, 139). However, this alternative methodological framing has substantial implications, and requires further justification.

² Wakefield also sometimes writes as if the judgments of Aristotle, Freud, and Foucault are as relevant as ours in determining the content of the contemporary concept of medical disorder (see Wakefield's exchange with Demazeux in this volume, and Wakefield 1999a). Do we have to contend as well with the judgments of "communities" past?

function plays no role in his “risky prediction,” as outlined in the above. All that is required for Wakefield’s thesis to be confirmed is that language users judge that something has “gone wrong inside a person” in a vague “organismic” sense. This point is helpfully made in the editors’ introduction, and Wakefield also leans on it throughout (see also Wakefield 1999b, 2000). One could, in principle, agree that language users think something has gone wrong inside a person when they judge that person to be diseased or disordered, while disagreeing that this “going wrong” amounts to a failure of evolved function. Wakefield’s conceptual analysis—the HDA—would, nonetheless, be confirmed by such an outcome.

So, with all these methodological caveats on the table, what is left of Wakefield’s risky bet? The community (understood vaguely) will only judge that some particular person has a medical disorder if they also believe (factoring in idiosyncratic beliefs and perceptions of the empirical facts) that there is something vaguely “going wrong” inside the person, and that this is bad. At this point, it seems to me that Wakefield’s much-discussed theory of mental disorders is in fact quite modest. All he is committed to is the view that people generally associate “having a medical disorder” with things going biologically wrong in a harmful way. Once clearly explicated, Wakefield’s HDA strikes me as extremely plausible and, indeed, unfairly derided. Who could possibly deny that a medical disorder has to do with bodily things going wrong in a bad way?

4. Wakefield’s “Black Box Essentialism”

Now that we have more of a handle on Wakefield’s conceptual analysis of medical disorder, let us move on to consider his supposedly nonconceptual claims about the nature of function. While Wakefield is a traditional conceptual analyst about “medical disorder”—making only contingent claims about the nature of our psychologies—he advocates “black box essentialism” about the concept of function (Wakefield 1999b, 2000). He is inspired here by the work of Hilary Putnam and Saul A. Kripke on natural kind terms but his application of this literature is not straightforward (see Putnam 1975; Kripke 1980). In what follows, I shall try to explain what I now take to be Wakefield’s view.

As noted, Wakefield advocates an evolutionary-selected effects theory of function. There are two well-known methodological approaches to the selected effects functions already present in the literature—that of Karen Neander and that of Ruth Garrett Millikan. Neander argues that we should view the selected effects theory as a conceptual analysis of the modern biologists’ notion of function (Neander 1991), while Millikan sees the selected effects theory as a description of the phenomenon itself, a real or theoretical definition (Millikan 1989, 2017). In Millikan’s view, concepts are externally individuated—indeed, the only publicly shared, temporally invariant feature of a concept is, generally speaking, its worldly referent (Millikan 2000). So, although Millikan is not a proponent of conceptual analysis, a description of what functions are like in the external world informs our understanding of the *concept*, in her view, because the identity of the concept does not depend on anything more than its referent.

Before reading *Defining Mental Disorder*, I thought Wakefield’s approach to selected effects functions was essentially that of Millikan. “Function” refers, rigidly, to naturally selected effect, and the selected effects theory is the right descriptive theory of these real-world phenomena. Assuming Millikanian content externalism about the concept of function, the content of the concept of function is, then, “naturally selected effect.” Thus,

“naturally selected effect” *does* describe the concept of function.³ I see now, however, that Wakefield’s position is, in fact, a more complex third way.

Contrary to Millikan, Wakefield posits a strict distinction between theories and concepts (see the exchange in this volume between Wakefield and Lemoine, as well as Wakefield 2000). Concepts are descriptions of what is in our heads; theories are descriptions of what is in the world. Therefore, facts about the kind to which “function” refers has nothing at all to do with the *concept*. The concept of function is, instead, a psychological criterion of the following form: *X* is a function if it is an instance with the same “essence” as certain psychological prototypes for biological function, such as sight. This is Wakefield’s conceptual claim about what is in our heads when we think about biological function and retains his broadly descriptivist internalism about conceptual content. This conceptual claim is then supplemented by the supposedly nonconceptual theoretical claim that the “essence” of biological function is “naturally selected effect.” This, for Wakefield, is strictly an empirical and scientific discovery about biological functions, and thus *not* a part of the concept.

There is some understandable confusion in the literature as to precisely how Wakefield’s conceptual analysis (the harmful dysfunction part) is supposed to relate to his evolutionary view of biological function and dysfunction. For example, in his chapter, Faucher sums up Wakefield’s view as follows:

One can thus understand that Wakefield is making two separate claims: (1) the correct analysis of our concept of mental disorder has to be made in terms of harm and dysfunction of a mental mechanism, and (2) the correct understanding of the concept of dysfunction is in terms of evolutionary function. (Faucher and Forest 2021, 49)

As I understand it, this is not really Wakefield’s view. The idea that functions are naturally selected effects is *not* a claim about the concept of biological function (because the referent is not part of the concept), but a nonconceptual scientific theory. So, there are really three moving parts in Wakefield’s overall view: (1) the psychological concept of mental disorder; (2) the psychological concept of function; and (3) the empirical, nonconceptual theory that the metaphysical essence of function is naturally selected effect.

Now that we better understand Wakefield’s black box essentialism about function, notice how philosophically substantial his commitments are in this regard. Once properly explicated, it seems that Wakefield’s thesis in fact rests on two distinct conceptual theses, one about mental disorder and one about function, as well a theoretical definition type-claim about the *nature* of functions (rooted in a theory of function as a natural kind). Usually, when people invoke the Putnam-Kripke externalism apparatus, it is to *avoid* having to do conceptual analysis in the traditional sense. However, Wakefield invokes these resources on top of a commitment to good old-fashioned conceptual analysis and internalism about conceptual content, making his view overall a strange hybrid of controversial methodological commitments. Charitably, one might say this makes Wakefield’s view nuanced and interestingly complex. Less charitably, one might say it makes it confusing and vulnerable to critique.

³ It should be noted that in recent work Millikan has abandoned the term “concept” in favor of the neologisms “unicept” and “unitracker” (Millikan 2017).

It seems to me that, perhaps, Wakefield's rather complicated black box essentialism about function is the "risky" part of his overall philosophical position and, arguably, the one we should really be debating—in contrast to his modest and plausible view of disorder as harmful dysfunction, which, upon clarification, I find it hard to disagree with. When it comes to Wakefield, have we been having the wrong debate? I shall close by offering three features of Wakefield's black box essentialism that, to my mind, are worthy of further examination.

Firstly, Wakefield talks as if definitions cannot be informed by science: "Millikan's 'theoretical definition' turns out not to be a definition at all but just a scientific discovery" (Wakefield 2000, 39). Elsewhere, he says: "The HD analysis cannot directly define disorder in evolutionary terms because the analysis aims to capture a widely shared, intuitive medical and lay concept that existed long before evolutionary theory" (1999a, 375). So, following Wakefield's reasoning here, it cannot be part of the concept of whale that whales are mammals, because there was a time when humans referred to whales and did not know that they were mammals. So "whale" must *mean* something else—maybe large, finned sea creature. There are two ways of responding to Wakefield here: one Neanderian; one Millikanian.

Neander would concede that perhaps "being mammals" was not part of "what we had in mind" in discriminating whales from non-whales prior to the discovery that whales were mammals but would also suggest that *now it is*. If our linguistic community of interest is contemporary biologists, arguably, they *do* now have this fact in mind (see Neander 1991). Wakefield seems oddly invested in the idea that the outcome of a conceptual analysis must transcend time in order to be adequate (see also his response to Demazeux in Faucher and Forest 2021). This strikes me as strange given that, on his view, a conceptual analysis is simply a contingent psychological claim about mental states that guides certain individuals in making classificatory judgments. Of course, these psychological criteria may change as science, culture, and knowledge progress. Why deny this?

Millikan, on the other hand, would remind us that, often, the only historically invariant part of a concept is its *referent*. As such, if a concept has a uniform referent throughout time, the stability of the concept is guaranteed by this alone (Millikan 2000, 2017; see also Sawyer 2020). So long as the referent remains the same, the concept is retained across time. There is, therefore, no need to insist, as Wakefield does, that we must have some shared *psychological state* in common with Aristotle in order to ensure conceptual continuity over time (see Wakefield 2000). We share a concept with Aristotle because we refer to the same *kind* of thing. Wakefield clearly denies this Millikanian picture but it is not clear on what grounds.

Second, Wakefield is committed to biological function being like a natural kind with a metaphysical "essence," where this essence is the process of "natural selection." I am myself not unsympathetic to this kind of claim but it is highly nontrivial. Usually, essences are thought of as properties that explain the non-accidental co-occurrence of *multiple* properties within a kind (see Godman, Mallozzi, and Papineau 2020). However, it is not clear to me that there is any "property clustering" in the case of biological functions (which properties would that be?), nor that natural selection *explains* the co-occurrence of properties in this sense. In short, it is by no means obviously plausible that "function" is a

natural kind term, nor that “natural selection” is the sort of property that could count as an essence.⁴

Third, Wakefield sometimes talks as if medical disorder is a black box essentialist concept (see Wakefield 1999b, as well as the exchange between Wakefield and Zachar in this volume). However, from my best understanding of Wakefield’s methodological commitments, disorder really *is not* a black box essentialist concept and it is really rather *important* for the coherence of his overall view that it is not. If mental disorder, like function, were a black box essentialist concept, all we could say about the “concept of disorder” is that disorders are whatever shares the underlying metaphysical essence of certain psychological prototypes of disease (say, cancer, cardiac arrest, Huntington’s disease). It would then be up to science, or perhaps scientifically informed philosophy, to make the further empirical discovery of what the underlying essence of medical disorder is. Such a picture leaves little room for the “harmful dysfunction” conceptual analysis of medical disorder on which Wakefield has staked his claim. So not only is disorder *not* a black box essentialist concept on Wakefield’s view, moreover it *cannot* be—on pain of Wakefield having to abandon the HDA.

A crucial part of Wakefield’s view, then, one that is left unsaid, is that medical disorder is a fundamentally different kind of concept to function. It is unclear, however—to me, at least—what makes this the case. Why isn’t “medical disorder,” like biological function, a black box essentialist concept?

In sum, *Defining Mental Disorder* is an extremely thought-provoking volume of work that serves to elucidate a number of under-discussed issues in Wakefield’s philosophy. The chapters contributed by critics provide a comprehensive taxonomy of a variety of philosophical issues raised by Wakefield’s work, while Wakefield himself performs admirably in defending and strengthening his influential view. I recommend it wholeheartedly to students and researchers in philosophy of medicine who draw on Wakefield’s analysis in their work, and to philosophically interested clinicians and empirical researchers who wish to gain a deeper understanding of underlying philosophical issues.

Acknowledgments

I would like to thank Justin Garson, Jerome Wakefield, and David Papineau for their helpful comments. I also gratefully acknowledge the financial support of the John Templeton Foundation (#62220). The opinions expressed in this paper are those of the author and not those of the John Templeton Foundation.

Disclosure Statement

No competing interest was reported by the author.

References

Boorse, Christopher. 1977. “Health as a Theoretical Concept.” *Philosophy of Science* 44, no. 4: 542–573. <https://doi.org/10.1086/288768>.

⁴ Wakefield is clear (see his responses to Zachar and Lemoine in this volume) that his notion of essence is very permissive, spanning traditional microstructural essences, as well as less traditional historical and functional “essences.” However, function is not obviously a historical or functional kind either, so further clarification is nevertheless in order.

- Deutsch, Max. 2021. "Conceptual Analysis without Concepts." *Synthese* 198, no. 11: 11125–11157. <https://doi.org/10.1007/s11229-020-02775-0>.
- Faucher, Luc and Denis Forest, eds. 2021. *Defining Mental Disorder: Jerome Wakefield and His Critics*. Cambridge, MA: MIT Press. <https://direct.mit.edu/books/book/5015/Defining-Mental-DisorderJerome-Wakefield-and-His>.
- Garson, Justin. 2019. "There Are No Ahistorical Theories of Function." *Philosophy of Science* 86, no. 5: 1146–1156. <https://doi.org/10.1086/705472>.
- Godman, Marion, Antonella Mallozzi, and David Papineau. 2020. "Essential Properties Are Super-Explanatory: Taming Metaphysical Modality." *Journal of the American Philosophical Association* 6, no. 3: 316–334. <http://dx.doi.org/10.1017/apa.2019.48>.
- Kripke, Saul A. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Machery, Edouard. 2009. *Doing Without Concepts*. Oxford: Oxford University Press.
- Millikan, Ruth Garrett. 1989. "In Defense of Proper Functions." *Philosophy of Science* 56, no. 2: 288–302. <https://doi.org/10.1086/289488>.
- . 2000. *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511613296>.
- . 2017. *Beyond Concepts: Unicepts, Language, and Natural Information*. Oxford: Oxford University Press.
- Neander, Karen. 1991. "Functions as Selected Effects: The Conceptual Analyst's Defense." *Philosophy of Science* 58, no. 2: 168–184. <https://doi.org/10.1086/289610>.
- Papineau, David. 2009. "The Poverty of Analysis." *Aristotelian Society Supplementary Volume* 83, no. 1: 1–30. <https://doi.org/10.1111/j.1467-8349.2009.00170.x>.
- Putnam, Hilary. 1975. "The Meaning of 'Meaning'." In *Mind, Language, and Reality*, Philosophical Papers vol. 2, 215–271. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511625251.014>.
- Sawyer, Sarah. 2020. "Truth and Objectivity in Conceptual Engineering." *Inquiry* 63, nos. 9–10: 1001–1022. <https://doi.org/10.1080/0020174X.2020.1805708>.
- Wakefield, J.C. 1992. "The Concept of Mental Disorder: On the Boundary between Biological Facts and Social Values." *American Psychologist* 47, no. 3: 373–388. <https://doi.org/10.1037//0003-066x.47.3.373>.
- . 1999a. "Evolutionary versus Prototype Analyses of the Concept of Disorder." *Journal of Abnormal Psychology* 108, no. 3: 374–399. <https://doi.org/10.1037//0021-843x.108.3.374>.
- . 1999b. "Mental Disorder as a Black Box Essentialist Concept." *Journal of Psychopathology and Clinical Science* 108, no. 3: 465–472. <https://doi.org/10.1037/0021-843X.108.3.465>.
- . 2000. "Aristotle as Sociobiologist: The 'Function of a Human Being' Argument, Black Box Essentialism, and the Concept of Mental Disorder." *Philosophy, Psychiatry & Psychology* 7, no. 1: 17–44. <https://muse.jhu.edu/article/28337>.
- . 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, no. 6: 648–682. <https://doi.org/10.1093/jmp/jhu038>.