Functional Indeterminacy, Addiction, and the Harmful Dysfunction Analysis

Abstract: According to Jerome Wakefield's harmful dysfunction account of mental disorder, a mental disorder must involve an objective dysfunction couched in evolutionary terms. However, selected effects functions are indeterminate, because the same trait can be both selectively advantageous and disadvantageous. Therefore, in some cases there may be a dysfunction, on the basis of which a psychiatric disorder is attributed, that can be described in multiple empirically adequate ways. The choices involved in these cases are value-laden. Some cases of addiction may fit this mold. Indeterminacy in the alternative descriptions of the states/processes/mechanisms involved in addiction implicates opposing value judgments.

Word Count: 4404

1. Introduction

Jerome Wakefield's harmful dysfunction account of mental disorder is widely influential in the philosophy of psychiatry (Wakefield 1992). The harmful dysfunction account analyzes the concept of mental disorder into two components: a descriptive component and a normative component. The descriptive component is cashed out in terms of the attribution of an objective psychological or physiological dysfunction couched in evolutionary terms, while the normative component is cashed out in terms of a value judgment to the effect that the dysfunction in question is disvalued according to local social norms. Wakefield's account is attractive for two reasons. First, it can employ the descriptive component to stave off anti-psychiatric claims (Szasz 1960) that psychiatrists use medicalized language to disingenuously sanction individuals that do not behave according to the local dominant social values. But Wakefield can also use the normative component to accommodate intuitions that social values play some role in the practice of psychiatry. The harmful dysfunction account would thus legitimate a science of mental disorder while prompting practitioners and patients to assess, debate, and negotiate social norms that come into play insofar as the physiological or psychological dysfunctions interact with our social practices.

My interest in this paper concerns the descriptive rather than the normative component of the harmful dysfunction account. Wakefield's view requires some notion of function/dysfunction such that all genuine mental disorders involve a dysfunction that can be identified on value-free, naturalistic terms. The attribution of the dysfunction must be value-free because the value-laden component of the disorder attribution is meant to be limited to a subsequent value judgment regarding the harm that the naturalistically-described dysfunction is causing. Wakefield's account

is predicated on the bet that all mental disorders can be neatly split into these two components, without any cross-contamination between them. However, I will argue, using addiction as a case study, that there are some cases in which a candidate-dysfunctional state, on the basis of which a disorder attribution might be made, can be legitimately described in multiple ways. These cases are at least conceptually possible, because the selected effects account of functions, which the harmful dysfunction account relies on, entails that functional attributions are indeterminate, since selective effects may have multiple empirically adequate descriptions. Relative to one description, a selective effect may be beneficial, but relative to another description, it may be disadvantageous. So a candidate-state might count as dysfunctional on one description, but not on others. In some cases, the choice of a description can make the difference between a candidate-state being described as dysfunctional or not being described as dysfunctional. Lastly, and most problematically for the harmful dysfunction account, it is likely that the choice of description is determined by implicit value judgments in these cases. If that is so, then there are at least some hard cases for which there is no naturalistic way to determine whether a state is dysfunctional or not. The general form of my argument is as follows:

1) Suppose the selected effects (SE) view is the correct view of function for psychiatric classification.[1]
2) Given that the same state can produce different effects on fitness in response to stimuli that can be described at different levels of grain in different environments throughout evolutionary history, there can be more than one empirically warranted way of characterizing the function of that state.

[1] Note that I am not committing to the SE view of functions itself here or that the SE view is the correct view of functions for psychiatric classification.

3) If this is true, then there is no single correct description of a state in terms of its (SE) functional or dysfunctional status, and thus in meeting the first condition required for counting as a mental disorder on Wakefield's view.
4) Values often determine the way in which theorists frame selective conditions, outcomes, or states/processes/mechanisms of relevance in these hard cases so as to yield determinate functions and/or contents.
5) Thus, there is no naturalistic (i.e., empirically decidable) way to determine whether that condition is functional or dysfunctional, and thus a mental disorder in some cases. The entire process is value-laden from the start.

I have decided to focus on addiction as a means of illustrating how functional indeterminacy can invite competing descriptions that invite different value judgments. I will argue that there are places where values enter: 1) in appeals to selective scenarios (e.g., the developmental-learning model of addiction), 2) in choices of cut-offs in the context threshold effects, given the multiple pathways by which addiction is promoted or sustained, and 3) in competing assessments of evolutionary mismatches.

2. The Indeterminacy of Selected Effects Functions

Since selected effects functions analyze the capacity in terms of its selection history, which is in turn given a naturalized rendering in terms of evolutionary theory, they are vulnerable to any of the problems that beset evolutionary thinking more broadly. Particularly in this case, the problem is that selected effects functions are indeterminate in a few related senses.

First, the same capacity can have both selective benefits and selective deficits. For example, the human capacity for abstract thought presumably had something to do with humans

coming to rule the natural world through sophisticated cooperation techniques, tool use, language, etc. (Sperber & Mercier 2017). At the same time, the capacity for abstract thought raises design quandaries that wouldn't have otherwise been present. A mind that can make inferences about the future based on past occurrences will readily come to the conclusion that personal death is inevitable. Plausibly, this inferential capacity can create a great deal of anxiety that no other animal has to deal with. Additionally, the use of folk psychology (the commonsense framework of beliefs, desires, intentions, etc.) that humans implement for mass cooperation also has its pitfalls. The meta-representational abilities that folk psychology enables also create the possibility of deception–getting the other to represent one's own intentional states inaccurately via some type of manipulative behavior. These are perennial adaptive problems that humans had to face in ancestral environments. One theory is that some forms of religion developed to combat these problems: for example, 1) concepts of the self as persisting beyond death may have evolved as a buffer against anxiety about inevitable death (Nichols et al. 2018), and 2) concepts of moralistic gods with the ability to police behavior via special knowledge capacities, including reading people's minds, may have developed in response to the problem of deception (Atran 2002, pg. 267-268).

Note that even in the examples I just mentioned, there are already tantalizing glimpses for how some psychiatric disorders may have developed. Anxiety disorders may be a sort of overfiring of capacities that underlie danger detection (Murphy 2005, pg. 747-752). Some delusional beliefs, including those with religious content, may result from an overfiring of folk psychological and/or more general social relationship capacities (sometimes known as the hyperactive agency detection device– Barrett 2000). Of course, all of these capacities were selected for design reasons (or "free-floating rationales; Dennett 1988), but the selection of those

capacities also enables the possibility of dysfunctions that may underlie psychiatric diagnoses. As just one more example in this vein, according to a prominent theory of depression, depression was adaptive in ancestral environments in response to a loss of status within the hunter-gatherer group (Price et al. 1994; see also Nesse & Williams 1995). Instead of fighting those with status and risking injury, abandonment, or death, it may have been adaptive for individuals that had just lost status to reevaluate their position in the dominance hierarchy and accept a role with lower-status within the group.

In each of these cases, we see how certain human mental capacities developed in response to adaptive problems in ancestral environments. But while the capacities that were selected for may have enhanced fitness overall, those same capacities may produce states that are not adaptive or lower fitness, as in the case of anxiety disorders, schizophrenic delusions, and depression, plausibly. Some of the unique cognitive capacities of human beings, like the ability to predict the future on the basis of the past and the ability to understand each other on a meta-representational level, also are disadvantageous in some respects. And it is here–reflection on examples such as these, that raises a problem for the view that mental dysfunctions can be objectively identified. The line between function and dysfunction is often blurry, and there may not be a no single correct description of a condition in terms of its (SE) functional or dysfunctional status, given that a description can either pick out a selective advantage or a selective deficit.

The first sort of indeterminacy that I discussed is only a form of indeterminacy in an informal or intuitive sense. However, the indeterminacy becomes formalized when it is understood from a teleosemantic perspective. Neander herself recognized that selected effects functions are indeterminate: “The problem is that for any given trait and any given function there

seems to be more than one way to describe that function, and conflicting judgments regarding biological norms can apparently be derived from the different descriptions" (Neander 1995, pg. 113). This complication arises because many traits/capacities/features have a role within a hierarchy of functions. Depending on the level of analysis, the function of the heart is to beat, pump blood, to distribute vital nutrients to other systems, to keep the organism alive, etc (see Garson 2014).

Similarly, Neander's example is one in which a trait of the antelope altered the structure of hemoglobin, caused higher oxygen intake, allowed the antelope to survive at higher altitude, and contributed to the antelope's fitness (Neander 1995, pg. 114-5). The problem is that there are several functional descriptions that are compatible with the feature's activity, since they capture the same extension (Neander 1995, pg. 120-121).

These same problems of functional indeterminacy may arise in some psychiatric cases. These cases are at least conceptually possible, even if they are not actual (as I think they might be, as in some cases of addiction). Tim Thornton argues that indeterminacy problems arise for any notion of psychiatric disorder that relies on a selected effects account of functions, since the psychiatric case is just a special instance of all indeterminacy problems that arise given an evolutionary account (Thornton 2021). Wakefield accepts that indeterminacy may be inherit to evolutionary explanations, but he does not think this poses any significant trouble for the harmful dysfunction account. Any indeterminacy in medicine or psychiatry can ultimately be traced back to indeterminacy in the background evolutionary biology, and medicine and psychiatry can inherit whatever solutions evolutionary biology presents (Wakefield 2021). The contention is that whatever indeterminacy in psychiatric explanations that depend on etiology are not especially problematic; that is "given evolutionary theory's determination of functions (however potential

indeterminacies are dealt with at that level), disorder can be defined with adequate determinacy from there" (Wakefield 2021).

I agree with Wakefield that whatever problems of indeterminacy that arise in etiological psychiatric explanation are unlikely to put psychiatry on a more shaky footing with respect to value-ladenness than somatic medicine. In fact, I'm willing to concede that problems of indeterminacy likely do arise in somatic medicine, and practitioners likely choose alternative functional descriptions based on value judgments in some cases. So my position that functional attributions in psychiatry are also often made on the basis of a value judgment is not meant to underwrite anti-psychiatric claims. Instead, my goal is to illuminate where values come into psychiatric practice so that these value judgments can be more easily discussed by practitioners.

3. Hard Cases of Addiction: Indeterminacy and Value Judgments

I now turn to addiction as a case study for values entering into the practice of attributing psychiatric disorder, based on the selection of alternative selected effect- (dys)functional descriptions, on the harmful dysfunction account. Below, I tentatively outline three ways in which this might happen. Note that in the case of addiction, and I imagine in the case of any other psychiatric disorder, value-ladenness could be realized in multiple ways. My argument centers on how natural selection may underdetermine functional descriptions, leaving further value judgments up to us, but there could plausibly be other ways for values to enter into the picture. Indeed, on a causal role account of functions, this is built into the perspectivalism of the view. Even with just evolutionary considerations, there are specific empirical hypotheses to consider for each psychiatric disorder, and any argument for value-ladenness from functional indeterminacy will require examining those hypotheses as would be appropriate. But here I

choose to focus just on three evolutionary considerations that support this argument in the case of addiction.

A) The Developmental-Learning Model of Addiction

The developmental-learning model of addiction provides support for the view that there is plenty of room for alternative empirically adequate interpretations of the mechanisms that underlie substance addiction. Marc Lewis (2017) makes this case most forcefully using both neural and behavioral levels of analysis. Proponents of the disease model of addiction use characteristic brain changes involving synaptic networks in the striatum (pursuit of rewards), amygdala (emotional regulation), hippocampus (memory encoding and retrieval), and the dorsolateral prefrontal cortex (reasoning, planning, self-control) to argue that addiction is a brain disease with neural signatures (8-9). The key concepts involved in evaluating this claim are self-organization and neuroplasticity. The brain is a self-organizing system in the sense that there is a "feedback loop between experience and brain change" that makes some mental states more probable to occur in the future than others (9-10). Eventually these processes lead to the development of behavioral habits. The brain is neuroplastic in the sense that as the "hardware" of the mind-brain system, it is designed to reconfigure in whatever way is necessary to sustain the changes that need to occur on the functional or intentional levels (10). So the challenge for proponents of the disease model of addiction is to provide reasons in favor of distinguishing addictive processes from normal neuroplastic and self-organizing brain processes. This, however, is not easy.

The brain changes a lot in the areas outlined above in response to objects, people, and situations that have a highly salient motivational significance. Plausibly, natural selection

selected those brain areas for the purpose of entrenching behavioral patterns and habits that would maximize the pursuit of those highly salient features of the world. These processes can also be described at the level of neurotransmitters, such as dopamine, which is particularly important for the brain changes that I am considering here. On some of the disease models of addiction, these brain processes are "co-opted" in cases of substance or behavioral addiction– the processes that were originally selected for more general pursuit of rewards have been employed for "rewards" that are not actually rewards. As Lewis (2017) says, "'addiction' doesn't fit a unique physiological stamp. It simply describes the repeated pursuit of highly attractive goals and the brain changes that condense this cycle of thought and behavior into a well-learned habit" (12). On what basis can addictive processes be distinguished from the mechanisms that underlie habit formation and pursuit of rewards or the automatization of behavior (and the underlying mechanism involving the growth of fibers from the VTA to the dorsal striatum)? This question becomes even more pressing when we look at behavioral addictions involving monetary rewards, romantic love/sexual partners, and so on, in which the line between "normal" and "disordered" behavior becomes even more blurred, since behaviors aimed at opportunity and safety (mediated through the cultural vehicle of money) and reproduction are likely to be directly promoted by natural selection.

So there is a degree of indeterminacy in the descriptions of the mechanisms that underlie addiction, on both a neural and behavioral level. Thus, in at least some cases, when describing those (token) states involved in the case of a putatively addicted individual, there are multiple empirically adequate ways to describe those states–some which will label those states as disordered, and others that won't. I conjecture that behavioral addictions involving, for example, Internet usage, pornography, cell phone usage, gambling, troubled love, etc. may be particularly

susceptible to multiple interpretations– natural selection may not always be able to tell us what was selected for and thus what is disordered. The issue cannot just be resolved by looking at the behavioral patterns and the underlying neural mechanisms, because those patterns and mechanisms are shared with normal non-pathological cases.

B) Threshold Effects

Values are also plausibly implicated in the diagnosis of addiction in determining whether a candidate disease state crosses a "threshold." An analogy can likely me made to cancer in this regard. The mechanisms underlying cancer are heterogenous and ambiguous enough that there are some cases in which there is no clear answer to whether the state is cancerous (Plutynski 2018, Chapter 2). While most cancers involve invasion, some cancer types such as sarcomas do not (73). In some cases, there are "borderline" tumors that have an "intermediate" malignant risk potential (Hageman, 2016). Cancers can originate through multiple pathways, and there may be a wide range of continuity between functional and dysfunctional states. Note that many other conditions share this same continuous feature, such as those that involve blood pressure or blood sugar level. Boorse recognizes this point: "The precise line between health and disease is usually academic, since most diseases involve functional deficits that are unusual by any reasonable standard" (1977, 559). This line-drawing problem is also connected to the problem of determining the appropriate reference class (age, sex, etc.) relative to which an assessment of normal function ought to be made (Kingma 2010). Plausibly, value judgments partially determine this assessment.

Similar threshold effects can also be found for psychiatric disorders, including addiction. First, there are disputes about whether addiction is a brain disorder, and if so, exactly neural

pathways are implicated in addiction (Wakefield, 2020). Addictions may, similar to cancers, be complex disease-entities that can involve a wide variety of cognitive mechanisms and brain states. Some authors argue that addictive behavior is caused by a stimulus-response mechanism that is entirely compulsive, either understood as automatic behaviors (Tiffany, 1990), or as behavior caused by very strong motivational states (Robinson & Berridge 1993). However, this interpretation of addiction is challenged by data that show that the addicted person sometimes chooses to remain abstinent if their background incentives are manipulated– for example, if they are offered prizes for continued abstinence (Silverman et al., 2016). This challenges both the notion that addicted individuals lack free will with respect to their addictive behaviors and the notion that addictive behavior is a matter of stimulus-response. Beliefs and desires with intentional content must be invoked to explain the addicted person's responses to these incentives, thus ruling out any hope for definitively demarcating addiction from non-addiction based on some signature of compulsion.

An opponent arguing that there are no threshold effects for addiction might then argue that there are distinct neural signatures that can disambiguate addictive from non-addictive states, even if the line cannot be clearly drawn based on stimulus-response mechanisms. However, this proposal is problematic for similar reasons to the ones I discussed above with respect to the developmental learning model of addiction. Brain changes themselves cannot distinguish between addictive and non-addictive states, and likewise there is no clear justified inference from mere brain difference to brain disease (Pickard 2021). The problem is to find some single brain signature that, for example, cases of alcohol addiction, drug addiction, and broader behavioral addictions share, but that other "addiction-like" states do not share. But without any such joint of nature that makes our nosological task easy and simple, we are left in a

situation analogous to the one we face in cancer. If so, there will be hard cases in which we have to make academic decisions with respect to a similar line-drawing problem. And if that is the case, then there will likewise be some value judgments as to whether someone is addicted– there might be an accumulation of risk factors and "addictive-like" brain changes, but no fact of the matter as to whether some mechanism is not performing its proper function– at least not yet. If there is no fact of the matter, our classificatory decision cannot, by definition, be naturalistically determined. If so, it must be determined by our values, at least in part.

C) Evolutionary Mismatches

The final way in which values may come into play in assessments of putative addictive states is competing assessments of evolutionary mismatches. An evolutionary mismatch comes about when the adaptations bequeathed by natural selection lag behind the rate of change of the environmental context of selection (Bourrat & Griffiths 2021). The genes that determine the phenotype of the organism may be adapted to maintaining fitness in an ancestral environment that no longer exists. In the new environment (after both temporal and spatial change), those genes may still operate as if the organism were in the ancestral environment, which may not be optimal for fitness. This occurs because those genes cannot "keep up" with the rate of change in the environment, leading to a mismatch. As an example, Type 2 diabetes is plausibly caused by a mismatch between genes that were designed for an ancestral environment in which it was adaptive to be "thrifty" in consuming food with a modern environment in which the same strategy has negative health consequences (Neel 1962).

How might addiction also be caused by an evolutionary mismatch? Well, one possible interpretation of at least some drug addictions is that normally-functioning biological

mechanisms are responding as designed, but are being exposed to novel quantities of certain stimuli (Nesse & Berridge 1997). Specifically, it is the fact that certain psychoactive drugs are now available purely and with direct routes of administration that ingesting such drugs can easily "hijack" brain mechanisms designed for the experience of pleasure and desire (Nesse & Berridge 1997). If this hypothesis is true, then addiction need not be caused by any tissue damage.

One might wonder why addiction being caused by an evolutionary mismatch might be a reason in favor of thinking that it might be indeterminate whether a putative addictive state is in fact a case of a dysfunction or not. The reason for thinking that this is so on views which employ the selected effects account of (dys)function (such as Wakefield's view) is that a disorder must involve a failure in the functioning of an "internal" mechanism on these views (Wakefield 1992, 240-241). This clause is introduced to distinguish genuine cases of disorders from disorder-like states that are socially disvalued (241). Wakefield's conclusion on this point is that "from the dysfunction perspective, the idea that the distress is intrinsic to the person's condition just means that the distress results from the failure of one of the individual's internal mechanisms to perform the function for which it was designed" (241).

Now, going back to the hijacking theory of addiction, it becomes unclear whether addiction would necessarily have to involve a dysfunction when this clause is taken into consideration. The hijacking theory of addiction posits that certain normally operating psychological mechanisms are "co-opted," but this description leaves the locus of the disturbance underdetermined. However, one plausible interpretation is that the psychological mechanisms involved (specifically the ones for pleasure and desire) are not malfunctioning, since it would only be useful for the novel substance to "co-opt" them if they were still functioning properly, so to speak. If so, the addictive state does not directly involve any

dysfunctions. Now, it is quite plausible that in some cases an addictive condition may eventually cause other mechanisms to malfunction, even to the point of death. But in that sort of case, the addictive condition would arguably only count as a risk factor for an actual disorder rather than a disorder in itself. I do not take it as obvious that this would be the correct interpretation, but I only want to point out that this would be a case in which it is contestable whether the condition involves a dysfunction or not in the relevant sense.

From here, there are two ways in which values may influence deciding between these two interpretations. First, socially constructed norms regarding the ingestion of addictive substances may influence the diagnosis of an observer. There might be norms in place regarding methamphetamine that have no corollary for caffeine, for example. Second, the extent to whether psychological mechanisms are "co-opted" enough or in the right manner to warrant labeling the condition as addiction and as a dysfunction/disorder may be an academic decision about threshold effects, collapsing a case like this into the ones discussed above.

4. Conclusion

I have argued that the selected effects account of function is a solid foundation for psychiatric nosology, but that despite the naturalistic stance that it affords us, there are still difficult cases in which value judgments make the difference with respect to demarcating normal mental activity and psychiatric pathology. Value-ladenness sneaks in the back door in these cases because functions can be indeterminate on a selected effects account of functions: there can be more than one empirically warranted way of characterizing the same condition as either/both "functional," and "dysfunctional." At this point, nature leaves us without an answer, and it is up to our value judgments– our framing of selective conditions, outcomes, or

states/processes/mechanisms– to determine whether a condition is dysfunctional and thus a disorder. In other words, the easy division between 1) value-free dysfunction attributions, and 2) value-laden disorder attributions on Wakefield's view fails in these hard cases.

Furthermore, I examined addiction as an example in which this type of value-ladenness may be present in hard cases. I argued that the general form of my argument may apply for addiction when we consider the developmental-learning model of addiction, threshold effects, and interpretations of evolutionary mismatch. Similar considerations may apply to other psychiatric disorders. If so, the selected effects account of functions may quell some naturalistic qualms about the foundations of psychiatry, but not all.

References

Atran, Scott. 2002. *In Gods We Trust: Exploring the Evolutionary Landscape of Religion*. Oxford:Oxford University Press.

Barrett, Justin L. 2000. "Exploring the natural foundations of religion." *Trends in Cognitive Sciences*, 4(1):29-34.

Boorse, Christopher 1977. "Health as a Theoretical Concept." *Philosophy of Science*, 44(4):542–573.

Bourrat, Pierrick, and Paul E. Griffiths. 2021. "The idea of mismatch in evolutionary medicine." *The British Journal for the Philosophy of Science*.

Dennett, Daniel C. 1988. "Evolution, Error and Intentionality." In: S*ourcebook on the Foundations of Artificial Intelligence*, eds. Derek Partridge and Yorick Wilks. Cambridge :Cambridge University Press.

Garson, Justin. 2014. "Why (a Form of) Function Indeterminacy is Still a Problem for Biomedicine, and How Seeing Functional Items as Components of Mechanisms Can Solve it."

Hageman, I. 2016. Personal Interview with Anya Plutynski. Washington University in St. Louis, St. Louis, MO.

Kingma, Elselijn. 2010. "Paracetamol, poison, and polio: Why Boorse's account of function fails to distinguish health and disease." *British Journal for the Philosophy of Science* 61(2):241-264.

Lewis, Marc. 2017. "Addiction and the Brain: Development, Not Disease." *Neuroethics* 10:7–18.

Murphy, Dominic. 2005. "Can evolution explain insanity?" *Biology and Philosophy* 20:745-766.

Neander, Karen. 1995. “Misrepresenting and Malfunctioning.” *Philosophical Studies* 79:109-141.

Neel, James V. 1962. “Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"?,” *American Journal of Human Genetics* 14(4):353-62.

Nesse, Randolph M. and George C. Williams. 1995. *Why We Get Sick*. New York: Times Books.

Nesse, Randolph M. and Kent C. Berridge. 1997. “Psychoactive drug use in evolutionary perspective.” *Science* 278(5335):63–66.

Nichols, Shaun, Nina Strohminger, Arun Rai, and Jay Garfield. 2018. “Death and the Self.” *Cognitive Science* 42:314-332.

Pickard, Hanna. 2021. “Is addiction a brain disease? A plea for agnosticism and heterogeneity.” *Psychopharmacology*, 1-15.

Plutynski, Anya. 2018. *Explaining Cancer: Finding Order in Disorder*. Oxford:Oxford University Press.

Price, John, Leon Sloman, Russell Gardner Jr., Paul Gilbert, and Peter Rohde. 1994. “The social competition hypothesis of depression.” *The British journal of psychiatry: the journal of mental science*, 164(3):309–315.

Robinson, Terry E., and Kent C. Berridge. 1993. “The neural basis of drug craving: an incentive-sensitization theory of addiction.” *Brain research reviews*, 18(3):247–291.

Silverman Kenneth, August F. Holtyn, and Reed Morrison. 2016. “The Therapeutic Utility of Employment in Treating Drug Addiction: Science to Application.” *Psychological Science* 2(2):203-212.

Sperber, Dan, and Hugo Mercier. 2017. *The Enigma of Reason: A New Theory of Human Understanding*, Penguin Books.

Szasz, Thomas S. 1960. "The myth of mental illness." *American Psychologist* 15:113-118.

Thornton, Tim. 2021. "Naturalism and Dysfunction." In *Defining Mental Disorder: Jerome Wakefield and His Critics*, eds. Luc Faucher and Denis Forest, 291-300. Cambridge, MA:MIT Press.

Tiffany, Stephen T. 1990. "A cognitive model of drug urges and drug-use behavior: Role of automatic and nonautomatic processes." *Psychological Review* 97(2):147–168.

Wakefield, Jerome. 1992. "Disorder As Harmful Dysfunction: A Conceptual Critique of DSM-III-R's Definition of Mental Disorder." *Psychological Review* 99:232–47.

Wakefield, Jerome. 2020. "Addiction from the harmful dysfunction perspective: How there can be a mental disorder in a normal brain." *Behavioural Brain Research* 389:112665.

Wakefield, Jerome. 2021. "Is Indeterminacy of Biological Function an Objection to the Harmful Dysfunction Analysis? Reply to Tim Thornton." In *Defining Mental Disorder: Jerome Wakefield and His Critics*, eds. Luc Faucher and Denis Forest, 301-306. Cambridge, MA:MIT Press.