*Philosophy of Science* (2025), **00**, 1–11
doi:10.1017/psa.2025.10150

PSA PHILOSOPHY OF SCIENCE ASSOCIATION
PHILOSOPHY OF SCIENCE

CONTRIBUTED PAPER

# Functional Indeterminacy, Addiction, and the Harmful Dysfunction Analysis

Matthew Kern

Department of Philosophy, Washington University in St. Louis, St. Louis, MO, USA
Email: kern.m@wustl.edu

(Received 25 January 2025; revised 25 March 2025; accepted 24 July 2025)

## Abstract

According to Jerome Wakefield's harmful dysfunction account, a mental disorder must involve an objective dysfunction couched in evolutionary terms. However, selected effects functions are indeterminate because (i) the same trait can be both selectively advantageous and disadvantageous, and (ii) the functional activity of a trait can be assessed according to conflicting norms, given the trait's place in a hierarchy of functions. Therefore, there may be a dysfunction 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, involving indeterminacy that invites opposing value judgments.

## 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 divides 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,[1] whereas the normative component is cashed out in terms of a judgment (according to local social norms) that the dysfunction in question has negative value, causing harm to the individual. Wakefield's account is attractive for two reasons. First, it can employ the descriptive component to stave off antipsychiatric claims (Szasz 1960, Foucault [1961] 2006) that psychiatrists use medicalized language to disingenuously sanction individuals who 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,

[1] For an argument that evolutionary theory is a promising place to look for a solution to the problem of the validity of psychiatric diagnosis, see Nesse and Jackson (2011).

© The Author(s), 2025. Published by Cambridge University Press on behalf of Philosophy of Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

debate, and negotiate social norms that come into play insofar as the physiological or psychological dysfunctions interact with our social practices.

My focus in this article concerns the descriptive rather than the normative component of the harmful dysfunction account. Wakefield's view requires 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, given that 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. Thus, a candidate-state might count as dysfunctional under one description but not under others. Furthermore, the choice of description is determined by implicit value judgments in these cases. It follows that there are at least some hard cases for which there is no naturalistic way to determine whether a state is dysfunctional or not.[2] 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.
2. Given that the same feature can produce different effects on fitness in response to stimuli that can be described at different levels of granularity in different environments throughout evolutionary history, there can be more than one empirically warranted way of characterizing the function of that feature.
3. If this is true, then there is no single correct description of a feature in terms of its (SE) functional or dysfunctional status, leaving it ambiguous as to whether the feature meets 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 feature's performance is functional or dysfunctional and thus (potentially) constitutes the basis of a mental disorder. The entire process is value-laden from the start.

[2] I do not develop a comprehensive account of mental disorder here, although I take the considerations I raise to favor a normativist view over a naturalist view. Alternatively, the view I favor could be described as a hybridist view because I still take there to be *some* empirical restrictions on admissible functional ascriptions. However, this hybridism would not be Wakefield's hybridism, where values are entirely sequestered from descriptive facts about dysfunctions.

I focus on addiction as a means of illustrating how functional indeterminacy can suggest competing descriptions that invite different value judgments. There are three places where values enter: (1) in appeals to selective scenarios (e.g., the developmental-learning model of addiction); (2) in choices of cutoffs in the context of 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

Because the SE view analyzes a feature in terms of its selection history, which is in turn given a naturalized rendering in terms of evolutionary theory, SE functions are vulnerable to the problems that beset evolutionary theory more broadly. SE functions are indeterminate in two related senses.

First, the same trait can have both selective benefits and selective deficits. For example, the human capacity for abstract thought presumably had something to do with the emergence of cooperation, tool use, and language—all of which were selectively advantageous (Sperber and Mercier 2017). At the same time, the capacity for abstract thought could carry with it selective disadvantages. Individuals who can make inferences about the future based on past occurrences can anticipate that personal death is inevitable. Plausibly, this inferential capacity can create a great deal of anxiety. 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 create the possibility of deception—getting the other to represent one's own intentional states inaccurately. 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 death (Nichols et al. 2018), and (2) concepts of moralistic gods with the ability to police behavior may have developed in response to the problem of deception (Atran 2002, 267–68).

The previous examples plausibly indicate how some psychiatric disorders may have developed. Anxiety disorders may be caused by an overfiring of capacities that underlie danger detection (Murphy 2005, 747–52). Some delusional beliefs, including those with religious content, may result from a misfiring of folk psychological and/or more general social relationship capacities (sometimes known as the *hyperactive agency detection device*; Barrett 2000). 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 and Williams 1995). Instead of fighting those with status and risking injury, abandonment, or death, it may have been adaptive for individuals who 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 human mental capacities may have developed in response to adaptive problems in ancestral environments. While the capacities that were selected for may have enhanced fitness overall, those same capacities may produce states that are not adaptive or result in lower fitness, as in the case of anxiety disorders, schizophrenic delusions, and depression. Some of the unique cognitive capacities of human beings, like the advanced ability to predict the future on the basis of the past and the ability to understand one another on a meta-representational level, are also disadvantageous in some respects. Reflection on examples such as these raises a problem for the view that mental dysfunctions can be objectively identified. There may not be a single correct description of a condition in terms of its (SE) functional or dysfunctional status, given that a description can pick out either a selective advantage or a selective deficit.

Neander, a defender of the SE account, or more precisely, the teleosemantic account of functions, recognized that SE 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, 113). She acknowledged that many traits have distinct roles within a hierarchy of functions. Depending on the level of analysis, the function of the heart is to beat, to pump blood, to distribute vital nutrients to other systems, to keep the organism alive, and so forth (see Garson 2014).[3] Because the trait's functional activity is nested within a hierarchy of functions, that activity can be assessed according to conflicting norms, producing several possibilities for what the proper function of the item is that are all compatible with the item's actual performances.[4] This is the second and more troublesome type of functional indeterminacy that a view of mental disorder based on SE functions faces.

Neander's biological example is one in which an alteration of hemoglobin structure in antelopes caused (a) higher oxygen intake, (b) allowing the antelope to survive at a higher altitude, thus (c) increasing relative fitness (Neander 1995, 114–15). The problem is that there are several (at least three) functional

[3] Garson (2019, chap. 7) and Fagerberg and Garson (forthcoming) have argued that SE proper functions are proximal functions and that so-called "distal" functions are beneficial effects associated with a trait performing its function. They argue that such functions can be made determinate on the basis of purely value-free mechanistic considerations, using the kind of functional analysis often associated with causal-role accounts (Cummins 1975; Craver 2013) but without the observer-dependency. Addressing this view would require an article of its own, but as Hundertmark and van den Bos (2024, 10) have pointed out, there is a range of fields, including evolutionary psychiatry, for which the ascription of distal *functions*, not just the acknowledgment of beneficial effects associated with a proximal function, is explanatorily necessary. Without this crucial premise that SE proper functions must be proximal functions, we are left in a situation where the functional description we designate as the proper function of some item (among the competing alternatives as given by the SE account) in a given context will be sensitive to our particular explanatory, medical, and social values. Two of Garson's (2019) arguments for the view gesture at a value that may be at play: We may want our functional ascriptions to be *informative* in the *context of medical intervention* (Garson 2019, 179). It is not clear that such considerations will always favor the proximal function as being the proper function, especially in a psychiatric context.

[4] I consider this type of indeterminacy to be a biological instantiation of the type of indeterminacy that Wittgenstein (1953) and Kripke (1982) identified on the basis of general considerations about rule-following. Thornton (2021), whom I cite later, makes the same connection.

descriptions that are compatible with the feature's activity because they capture the same extension (Neander 1995, 120–21).

These same problems of functional indeterminacy may arise in psychiatric cases. These cases are at least conceptually possible. Thornton (2021) argues that indeterminacy problems arise for any notion of psychiatric disorder that relies on an SE account because the psychiatric case is just a special instance of all indeterminacy problems that arise given an evolutionary account. Wakefield accepts that indeterminacy may be inherent to evolutionary explanations, but he does not think this poses any significant trouble for the harmful dysfunction account. He argues that any indeterminacy in medicine or psychiatry can ultimately be traced back to indeterminacy in evolutionary biology, and medicine and psychiatry can inherit whatever solutions evolutionary biology presents (Wakefield 2021). The contention is that whatever indeterminacy there is in psychiatric explanations that depend on etiology is not uniquely 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, pg. 586).

I agree with Wakefield that whatever problems of indeterminacy that arise in psychiatric explanation are unlikely to put psychiatry on a shakier footing with respect to value-ladenness than somatic medicine. In fact, I'm willing to concede that problems of indeterminacy do arise in somatic medicine. There are many cases in which organisms face evolutionary trade-offs in fitness[5]—for example, between traits that may promote reproductive fitness but compromise survival. In such cases, practitioners choose between alternative functional descriptions based on value judgments. That functional attributions in psychiatry are also often made on the basis of value judgments is not meant to underwrite antipsychiatric 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, without being obscured by debates about objective dysfunctions.

## 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 SE (dys) functional descriptions. I outline three ways in which this might happen. Note that in the case of addiction, value-ladenness could be realized in multiple ways, apart from how natural selection may underdetermine functional descriptions. Indeed, on a causal role account of functions, this is built into the perspectivalism of the view (Craver 2013). On the SE account of functions, there are underdetermination problems concerning specific empirical hypotheses to consider for each psychiatric disorder, and any argument for value-ladenness from functional indeterminacy will require examining those hypotheses. But here, I choose to focus on just three considerations that support this argument in the case of addiction.

[5] Two types of such trade-offs are antagonistic pleiotropy and antagonistic coevolution (Crespi and Summers 2005). Thanks to Anya Plutynski for this point.

### 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 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 dorsolateral prefrontal cortex (reasoning, planning, self-control) to argue that addiction is a brain disease with neural signatures (Lewis 2017, 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 (Lewis 2017, 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 (Lewis 2017, 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 just outlined 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"—that is, 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) puts it, "'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, the pursuit of rewards, or the automatization of behavior more generally? 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 "dysfunctional" 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 the neural and behavioral levels. Thus, in at least some cases, when describing those (token) states, there are multiple empirically adequate ways to describe them—some of which will result in labeling those states as dysfunctional and others that won't. I conjecture that behavioral addictions involving, for example, internet usage, pornography, cell phone usage, gambling, troubled love,

and so forth 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 dysfunctional. The issue cannot be resolved by looking at the behavioral patterns and the underlying neural mechanisms because those patterns and mechanisms are shared with normal, nonpathological 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 be made to cancer in this regard. There are some cases in which there is no clear answer as to whether a state is cancerous (Plutynski 2018, chap. 2). For instance, there are "borderline" tumors that have an "intermediate" malignant risk potential (Ian Hagemann, personal interview with Anya Plutynski, December, 12, 2016). 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 in the context of his theory, but it is also relevant here: "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).

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 what 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, understood either as automatic behavior (Tiffany 1990) or as behavior caused by very strong motivational states (Robinson and Berridge 1993). However, this interpretation of addiction is challenged by data showing 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 and 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 nonaddiction 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 nonaddictive states, even if the line cannot be clearly drawn based on stimulus-response mechanisms. However, this proposal is problematic for reasons similar to the ones I discussed earlier with respect to the developmental-learning model of addiction. Brain changes themselves cannot distinguish between addictive and nonaddictive states, and likewise, there is no sound inference from mere brain difference to brain disease not mediated by values (Pickard 2022). 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. Without any such joint in nature, 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" (conventional) decisions. And if that is the case, then there will likewise be some value judgments as to whether someone is addicted. In any given case of a prospective addict, there might be an accumulation of risk factors and "addictive-like" brain changes, but drawing the line is not possible via appeal to an unambiguous failure of function of some mechanism—at least not yet. If there is no fact of the matter, our classificatory decision cannot, by definition, be naturalistically determined.

### *C. Evolutionary mismatches*

The final way in which values may come into play in assessments of putative addictive states is through competing assessments of evolutionary mismatches. An evolutionary mismatch comes about when adaptations bequeathed by natural selection lag behind the rate of environmental change (Bourrat and Griffiths 2024). The genes that determine the phenotype of the organism may be adapted to maintaining fitness in an ancestral environment. In the new environment (after both temporal and spatial change), those genes may still operate as if the organism were in the ancestral environment, leading to a mismatch. For example, common diseases in the postindustrial era, such as type 2 diabetes and heart disease, are plausibly caused by an interaction between the effects of our modern sedentary lifestyles and genes that were designed to assist nomadic humans with their more physically active lifestyles (Lieberman 2015).

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 they are being exposed to novel quantities of certain stimuli (Nesse and Berridge 1997). Specifically, certain psychoactive drugs are now available that can easily "hijack" brain mechanisms designed for the experience of pleasure and desire (Nesse and 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 is a reason in favor of thinking that it is indeterminate whether a putative addictive state is in fact a case of a dysfunction. The reason for thinking that this is so on Wakefield's view is that a disorder must involve a failure in the functioning of an "internal" mechanism on his view (Wakefield 1992, 240–41). This clause is introduced to distinguish genuine cases of disorders from disorder-like states that are merely socially disvalued (Wakefield 1992, 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" (1992, 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 because 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 dysfunction(s) rather than a dysfunction in itself. I do not take it as obvious that this would be the correct interpretation; I only want to point out that this would be a case in which it is contestable whether the condition involves a dysfunction in the relevant sense.

From here, there are two ways in which values may influence a decision 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 which psychological mechanisms are "co-opted" in the right manner to warrant labeling the condition as addiction and as the result of a dysfunction may be an academic decision about threshold effects, collapsing a case like this into the ones discussed earlier.

## 4. Conclusion

I have argued that although the SE account of function may appear to provide a plausible naturalistic foundation for psychiatric nosology, there are still difficult cases in which value judgments are required to demarcate normal mental activity and psychiatric pathology. Value-ladenness sneaks in the back door in these cases because functions can be indeterminate in an SE account: There can be more than one empirically warranted way of characterizing the same condition as either/both "functional" and/or "dysfunctional." At this point, nature leaves us without an answer. It is our value judgments—our framing of selective conditions, outcomes, or states/processes/mechanisms—that 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. The general form of my argument may apply to addiction when we consider the developmental-learning model of addiction, threshold effects, and variable interpretations of evolutionary mismatch. Similar considerations may apply to other psychiatric disorders. If so, the SE account of functions may quell some naturalistic qualms about the foundations of psychiatry, but not all.[6]

[6] I would like to thank Anya Plutynski, Carl Craver, Justin Garson, and Bennett Knox, among many others, as well as audiences at the 2024 meetings of the Southern Society of Philosophy and Psychology, the Society for Philosophy and Psychology (poster session), and the Philosophy of Science Association, as well as at the 2025 meeting of the Eastern Division of the American Philosophical Association, for their feedback on earlier drafts of this article. I would also like to thank my colleagues at Washington University in St. Louis who provided feedback on an earlier draft at our Work-in-Progress Series in March 2024. Any errors that remain are my own.

## 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. https://doi.org/10.1016/S1364-6613(99)01419-9.

Boorse, Christopher. 1977. "Health as a Theoretical Concept." *Philosophy of Science* 44 (4):542–73. https://doi.org/10.1086/288768.

Bourrat, Pierrick, and Paul Griffiths. 2024. "The Idea of Mismatch in Evolutionary Medicine." *British Journal for the Philosophy of Science* 75 (4):921–46. https://doi.org/10.1086/716543.

Craver, Carl F. 2013. "Functions and Mechanisms: A Perspectivalist View." In *Functions: Selection and Mechanisms*, edited by Philippe Huneman, 133–58. New York: Springer.

Crespi, Bernard, and Kyle Summers. 2005. "Evolutionary Biology of Cancer." *Trends in Ecology & Evolution* 20 (10):545–52. https://doi.org/10.1016/j.tree.2005.07.007.

Cummins, R. 1975. "Functional Analysis." *Journal of Philosophy* 72 (20):741–64. https://doi.org/10.2307/2024640.

Dennett, Daniel C. 1988. "Evolution, Error and Intentionality." In *Sourcebook on the Foundations of Artificial Intelligence*, edited by Derek Partridge and Yorick Wilks, 190–211. Cambridge: Cambridge University Press.

Fagerberg, Harriet, and Garson, Justin. Forthcoming. "Proper Functions Are Proximal Functions." *British Journal for the Philosophy of Science*. https://doi.org/10.1086/731869.

Foucault, Michel. (1961) 2006. *Madness and Civilization*. New York: Vintage Books.

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." PhilSci-Archive. https://philsci-archive.pitt.edu/10899/

Garson, Justin. 2019. *What Biological Functions Are and Why They Matter*. New York: Cambridge University Press. https://doi.org/10.1017/9781108560764

Hundertmark, Fabian, and Marlene van den Bos. 2024. "Biological Functions and Dysfunctions: A Selected Dispositions Approach." *Biology and Philosophy* 39 (2):1–20. https://doi.org/10.1007/s10539-024-09944-2.

Kripke, Saul A. 1982. *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.

Lewis, Marc. 2017. "Addiction and the Brain: Development, Not Disease." *Neuroethics* 10 (1):7–18. https://doi.org/10.1007/s12152-016-9293-4.

Lieberman, Daniel E. 2015. "Is Exercise Really Medicine? An Evolutionary Perspective." *Current Sports Medicine Reports* 14 (4):313–19. https://doi/10.1249/JSR.0000000000000168.

Murphy, Dominic. 2005. "Can Evolution Explain Insanity?" *Biology and Philosophy* 20:745–66. https://doi.org/10.1007/s10539-004-2279-3.

Neander, Karen. 1995. "Misrepresenting and Malfunctioning." *Philosophical Studies* 79:109–41. https://doi.org/10.1007/BF00989706.

Nesse, Randolph M., and Kent C. Berridge. 1997. "Psychoactive Drug Use in Evolutionary Perspective." *Science* 278(5335):63–66. https://doi.org/10.1126/science.278.5335.6.

Nesse, Randolph M., and Eric D. Jackson. 2011. "Evolutionary Foundations for Psychiatric Diagnosis: Making DSM-V Valid." In *Maladapting Minds: Philosophy, Psychiatry, and Evolutionary Theory*, edited by Pieter R. Adriaens and Andreas De Block, 167–91. Oxford: Oxford University Press.

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

Nichols, Shaun, Nina Strohminger, Arun Rai, and Jay Garfield. 2018. "Death and the Self." *Cognitive Science* 42 (suppl. 1):314–32. https://doi.org/10.1111/cogs.12590.

Pickard, Hanna. 2022. "Is Addiction a Brain Disease? A Plea for Agnosticism and Heterogeneity." *Psychopharmacology* 239 (4):993–1007. https://doi.org/10.1007/s00213-021-06013-4.

Plutynski, Anya. 2018. *Explaining Cancer: Finding Order in Disorder*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780199967452.001.0001

Price, John, Leon Sloman, Russell Gardner Jr., Paul Gilbert, and Peter Rohde. 1994. "The Social Competition Hypothesis of Depression." *British Journal of Psychiatry* 164 (3):309–15. https://doi.org/10.1192/bjp.164.3.309.

Robinson, Terry E., and Kent C. Berridge. 1993. "The Neural Basis of Drug Craving: An Incentive-Sensitization Theory of Addiction." *Brain Research: Brain Research Reviews* 18 (3):247–91. https://doi.org/10.1016/0165-0173(93)90013-p.

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):203–12. https://doi.org/10.1037/tps0000061.

Sperber, Dan, and Hugo Mercier. 2017. *The Enigma of Reason: A New Theory of Human Understanding*. London: Penguin Books. https://doi.org/10.4159/9780674977860.

Szasz, Thomas S. 1960. "The Myth of Mental Illness." *American Psychologist* 15 (2):113–18. https://doi.org/10.1037/h0046535.

Thornton, Tim. 2021. "Naturalism and Dysfunction." In *Defining Mental Disorder: Jerome Wakefield and His Critics*, edited by Luc Faucher and Denis Forest, 291–300. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9949.003.0028.

Tiffany, Stephen T. 1990. "A Cognitive Model of Drug Urges and Drug-Use Behavior: Role of Automatic and Nonautomatic Processes." *Psychological Review* 97 (2):147–68. https://doi.org/10.1037/0033-295X.97.2.147.

Wakefield, Jerome. 1992. "Disorder as Harmful Dysfunction: A Conceptual Critique of DSM-III-R's Definition of Mental Disorder." *Psychological Review* 99 (2):232–47. https://doi.org/10.1037/0033-295X.99.2.232.

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. https://doi.org/10.1016/j.bbr.2020.112665.

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*, edited by Luc Faucher and Denis Forest, 301–6. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9949.003.0029.

Wittgenstein, L. 1953. *Philosophical Investigations*. New York: Wiley-Blackwell.

**Cite this article:** Kern, Matthew. 2025. "Functional Indeterminacy, Addiction, and the Harmful Dysfunction Analysis." *Philosophy of Science*. https://doi.org/10.1017/psa.2025.10150