15 Do the Works of Carl Craver or Marcel Weber Explain How Causal Role Functions Can Provide Objective Medical Judgments of Dysfunction? Supplementary Reply to Dominic Murphy

Jerome Wakefield

We have seen that Murphy fails to make his case that there is a way that CR functions can explain the objectivity of medical dysfunction in a way comparable to the objective account of the harmful dysfunction analysis (HDA; see the main reply to Murphy for references). However, he cites the works of Carl Craver and Marcel Weber as providing such an explanation. In this supplement to my reply to Murphy, I examine whether Craver's or Weber's work does provide such an explanation.

Carl Craver's (2007) Analysis of Mechanistic Explanation

Murphy prominently cites Carl Craver as a philosopher whose work can provide an explanation for how, despite interest relativity, CR functions can nonetheless provide the basis for an objective naturalist account of biological functions and thus a foundation for medical diagnostic judgments. In evaluating Murphy's claim, I will consider the book by Craver (2007) that Murphy cites as well as a related article by Craver (2001) that Craver cites in his book. I shall argue that Craver is not addressing, let alone solving, the problem of objectivity confronting Murphy's "function" claim. I hasten to add that this is not a criticism of Craver. Craver's work is precise, insightful, and illuminating regarding the problem he does address, which is the nature of an adequate CR mechanistic explanation.

Working within a broadly Cummins-inspired CR-function framework, Craver (2007) elaborates a "causal-mechanical model of constitutive explanation" in neuroscience—or simply "mechanistic explanation" (2007, 107): "Mechanistic explanations are constitutive or componential explanations: they explain the behavior of the mechanism as a whole in terms of the organized activities and interactions of its components" (2007, 128). His aim is to show how CR analysis "will have to be amended and revised if it is to offer a normatively adequate account of constitutive mechanistic explanation" (2007, 107). Explanatory adequacy is defined in comprehensive terms: "The central criterion of adequacy for a mechanistic explanation is that it should account for the multiple features of the phenomenon, including its precipitating conditions, manifestations,

inhibiting conditions, modulating conditions, and nonstandard conditions" (2007, 139). In Craver's CR-inspired model, multilevel hierarchical, contextual, and etiological causal processes in the component parts of a mechanism combine to explain the full set of dispositional capacities conferred on the containing mechanism.

He thus embraces Wesley Salmon's (1984, 1998) causal-mechanical analysis of explanation, rejecting covering-law, unificationist, and prototype-activation accounts. He also rejects reductionist demands to formulate theories at one privileged ontological level, instead allowing multiple mechanism levels to enter into adequate explanations. Unlike Cummins, Craver's exclusive interest is in explanations of a system's capacities in terms of its components at various sublevels. In contrast, Cummins's account of functional analysis allows noncomponential causal stories as well, such as psychological explanations in which a capacity is analyzed in terms of other capacities at the same level (e.g., the capacity to bake a cake in terms of the capacity to look up the recipe, to follow directions, etc.; or the capacity for action in terms of the capacity to have belief and desire reasons for the action).

Now, with respect to Murphy's claims, Craver acknowledges what Murphy will not accept, that the essential interest relativity of CR functions means that mechanistic explanation applies equally to healthy and disordered conditions. Thus, CR functions offer no resources to make this foundational medical distinction in any objective way:

> If one is interested in explaining the behavior of a mechanism under diseased or industrial conditions, then one will be interested in componency relations under those conditions. Although we are often interested in states of health or features that have been selected for, there is no reason to insist upon this restriction. There is no way to know what constitutes the "appropriate" conditions without specifying the pragmatic context in which one is operating. (Craver 2007, 155–156)

Craver is quite explicit that functions in the CR sense can be purely destructive and pathological:

> My account of mechanistic role functions does not appeal to any sense of adaptiveness in an environment; instead it appeals only to roles in contextual systems. These contextual systems may be adaptive or destructive, and they need not even be the kinds of systems for which talk of adaptation is appropriate. Heart disease, high blood pressure, cardiac arrhythmia, and arterial hardening all have mechanisms that span multiple levels, and this three-tiered perspective is as useful in those contexts as in those that are adaptive. Descriptions of hierarchical mechanisms are always descriptions of the mechanisms for … something that one wants to understand (build, control, predict) … without necessarily being adaptive or maladaptive. (Craver 2001, 67)

Sometimes Craver seems for a moment to write as if there is an objective and naturalist CR distinction between health and pathology, but he then steps back and clarifies that the CR view rejects any such notion due to the relativization of function to investigator interests and values:

> Finally, the component should be physiologically plausible. It should not exist only under highly contrived laboratory conditions or in otherwise pathological states. Of course, what constitutes a contrived condition or a pathological state varies across explanatory contexts. If one is trying to explain healthy functions, then pathological conditions might be considered physiologically implausible. If, on the other hand, one is trying to explain a disease process, one's explanation might be physiologically implausible if it assumes conditions only present in healthy organisms. What matters is that the parts' existence should be demonstrable under the conditions relevant to the given request for explanation of the phenomenon. (2007, 132)

In keeping with these views, Craver frequently provides examples of CR-type analysis in which the "functions" will be the production of a medically pathological condition: "A broken … kidney's mechanistic role is then identified against the fixed backdrop of a description of the way the circulatory system generally works, or the way that it preferably works, or the way that it works in whatever (normal or pathological) mechanism that we seek to understand" (Craver 2001, 72): "We provide an etiological explanation of why John is a victim of heart disease when we blame his smoking and diet and, perhaps, the mechanisms by which smoking and diet produce heart disease" (2001, 69–70); "Some mechanistic explanations … explain an event by describing its antecedent causes. Dehydration is part of the etiological explanation of thirst. Prion proteins are part of the etiological explanation of Creutzfeldt-Jacob disease. Excessive repetition of the CAG nucleotide pattern on the fourth chromosome is part of the etiological explanation for Huntington's disease" (2007, 107–108).

Given the interest relativity of the CR-mechanistic explanation, Craver conceptualizes an organ being "broken" as a deviation in how it works from any standard or class of entities one chooses. He rejects the statistical approach to brokenness because some CR functions concern statistically rare or manufactured phenomena. The baseline for judging what is broken can be what is statistically common, or what we prefer, or even a pathological trajectory in which we are interested. Thus, if one is studying the cause of progressive renal failure, a "broken" kidney might be one that spontaneously begins to function normally.

To the objection that one can define health within the mechanistic approach by simply identifying the organism's standard functioning, Craver makes clear that, given the interest relativity of all such analyses, what is normal or standard is not an independently definable notion but itself relative to the interests of the investigator. Sometimes that determining context might be the health of the organism, but that is an additional extra-CR concept imposed by the investigator and not defined by the functional analysis:

> "Normal" and "standard" conditions amount to something like "the way that the mechanism behaves under the conditions that we consider most appropriate for our current explanatory purposes." Sometimes this is assessed in terms of the healthy and fit organism, and normal means something like "behavior consistent with or conducive to overall system health and

> function." Sometimes it is assessed in terms of evolutionary stories, and so means something like "behavior similar to that which preserved the trait in the population of organisms." Sometimes normalcy is assessed in terms of its utility for an experiment, and so means something like "behavior consistent with or conducive to manipulation and detection with my experimental protocol." There is no need to be more restrictive about this notion. "Normal" and "standard" are defined relative to an implied investigative context. (Craver 2007, n. 13, 127)

Given such choices due to the interest relativity of CR functions, Craver alludes to the inevitable issue of value intrusion and bias that can enter into CR function judgments:

> Describing an item's mechanistic role is a perspectival affair. This perspectival take on functional ascription should be a reminder that what we take as functional descriptions can be tinged in a very direct way by our interests and biases. (Craver 2001, 73)

In the context of psychiatry, such perspectival neutrality on what is normal versus pathological can yield rationalization of social control of deviance. That is, like all Cummins-inspired views, Craver's analysis allows social values to determine what is of interest and thus to dictate functional hierarchies. This provides no protection from what Murphy portrays as the objection to diagnosis of the medical skeptic, who says that disorder judgments are just based on social value criteria. This approach is dangerously confusing when it comes to psychiatric diagnostic judgments of function and dysfunction. Suddenly, to use some standard examples, the interests of slaveholders in the antebellum South can allow for a legitimate "dysfunction" attribution to runaway slaves (the infamous diagnosis of "drapetomania"), and from the Soviet state's perspective, the Soviet dissidents can be legitimately considered to suffer from a mental dysfunction. We know that these are in fact illegitimate diagnoses because the "dysfunction" judgments underlying them are *false*, but that is a conclusion that cannot be reached objectively from a CR starting point that sees all capacities of the containing system as equally potential functions depending on the interests of the investigator. Relativizing notions of biological function and dysfunction is an assault on the legitimacy of disciplines like psychiatry that are already prone to confuse social undesirability and deviance with mental disorder. The objectivity of the function/dysfunction distinction makes medicine special and allows it to be more than a servant of social control. In effect, the CR approach yields a normative approach to disorder attribution by way of perspectival definition of function and dysfunction, yielding to Murphy's skeptic who says diagnosis is just about values. Murphy's attempt to reject the skeptic's position is undermined by his embrace of CR functions.

However, against some evolutionists, Craver obviously rejects the notion that SE functions aimed at explaining normality are the only kind of real functions. He insists on the need for CR mechanistic analyses that can be equally aimed at disordered or other anomalous features not amenable to SE explanation:

> Some workers in the systems tradition assume or stipulate that all explanandum phenomena have been selected by evolution by natural selection or that the phenomena are otherwise adaptive (that is, the phenomenon is how something behaves when it is behaving properly). In the philosophy of biology, Cummins is best known for his attacks on Wright's adaptive view of functions. I side with Cummins. Neuroscientific explanations often focus on malfunctions, disease states, laboratory phenomena, pharmaceutical contrivances, and industrial and military applications (for example, how the vestibular system works in zero-gravity). … No doubt, some of the features of the brain have straightforward adaptive etiologies, but I do not want to presuppose for present purposes that all of them do. Either way, one still needs the more limited sense of role-functions, activities that make some crucial contribution to the behavior of a containing system. (2007, n. 10, 124)

In rejecting the claim that all neurobiological explanatory activity in every area should be in terms of SE functions, Craver here expresses a pluralism in which there are explanatorily rich SE functional explanations needed in some areas but where of course "more limited" CR functions are also needed. Note that he suggests that SE explanations are especially appropriate when one is interested in "how something behaves when it is behaving properly" (the medical notion of proper function or health) but that CR functions are especially needed to study departures from biologically designed functioning such as "malfunctions, disease states, laboratory phenomena, pharmaceutical contrivances, and industrial and military applications (for example, how the vestibular system works in zero-gravity)." I would say that in the end, both CR and SE analysis is necessary to understand both domains, but in particular that the disorder/nondisorder part of "behaving properly" in the objective adaptationist more-than-perspectival-interest sense is strictly and essentially dependent on SE functional analysis. Everything Craver says is consistent with this understanding.

In sum, Craver's work does not in any way support Murphy's view that CR functions can be naturalized in such a way as to support objective judgments of disorder and pathology. To the contrary, Craver is crystal clear that health and disorder are equally mechanistically explainable, are both of interest, and thus, relative to the mechanisms that produce them, are equally "functions" in the CR sense. He offers no account of how CR functions can objectively distinguish health from disorder and even implies that SE functions may well be needed for such a discrimination. Thus, he provides Murphy no lifeboat in which to escape the objections to his claim that CR functions are an adequate foundation for psychiatric diagnosis, given that Murphy accepts that disorder attributions must be objective judgments that something has gone wrong "under the hood" of the individual. Consequently, Craver's analysis offers no support for Murphy's contention that philosophers are resolving the problem of how CR functions can enable one to draw an objective naturalistic medical distinction between health and pathology.

Marcel Weber's Coherence Theory of Natural Function

I have been considering Murphy's response to a key objection to using CR functions as a basis for attribution of psychiatric and medical disorder. The objection, which he acknowledges as a serious one, is that CR functions, due to their being interest relative, cannot explain the objectivity of biological function and dysfunction as these concepts are applied in medical diagnosis. His response is to cite other philosophers who, he says, are formulating ways in which to construct relevant objective function judgments from CR functions. We have seen in my reply to Murphy and above that his citations of Bock and von Wahlert (1965), Amundson and Lauder (1994), and Craver do not support his point because they are pluralists about "function's" meanings who are not trying to do what Murphy describes.

However, Murphy's fourth and final citation to Marcel Weber's work on functional organization offers more hope for Murphy's point. Weber, in the course of a larger discussion of holism, explicitly frames his work in part as an attempt to explain how a CR-function approach can support a naturalist, objective account of biological function. I now consider whether Weber's approach succeeds in vindicating the ability of CR functions to explain biological functions without the addition of SE functions. Murphy cites an article (Weber 2005a) and a book (Weber 2005b) by Weber as supporting his point. My discussion is based on both, but I primarily focus on the article because Weber says that in it, he presents a "modified version" of his similar remarks in the relevant pages (35–39) of the book.

The main theme of Weber's article is, as Murphy suggests, to try to present an objective naturalistic account of biological functions within the constraints of the CR analysis of function. (Weber prefers to refer to CR functions as "Cummins functions" or as the "dispositional" concept of functions, and I use all three terms interchangeably here.) The challenge, as we have seen, is that CR functions are interest-relative and thus neither objective nor naturalist. Weber fully accepts this challenging starting point, characterizing Cummins's account of function as follows: "X's function in system S is φ exactly if X's capacity to φ is part of an adequate analytic account of S's capacity to ψ" (190). The choice of ψ is entirely unconstrained by this formula and depends on the interests of the investigator. Consequently, the fact that X's capacity to φ is part of an account of the outcome of interest ψ is sufficient for concluding that φ is a function of X within the system. Weber is also quite clear on how this CR approach to functions differs fundamentally from the SE approach: "What is crucial with this account is that function ascriptions according to this definition do not explain the *presence* of the function bearer in the system. In other words, the identification of something as a function entails nothing about why this thing is part of the system. In contrast, the etiological account of functions holds that this is precisely what a functional ascription explains" (190).

Given that any organismic capacity, healthy or disordered, can be explained causally by reference to the organism's parts, Weber acknowledges that the interest relativity of CR functions implies a prima facie mismatch between the CR account of function and a standard usage of "function" within certain areas of biology in which a function is not relative to the researcher's interests but an objective fact about the organism:

> Cummins … fully accepts the consequence that, on his account, the overall systems capacity is ours to choose, and it does not appear to be among his goals to naturalize functions. However, it seems to be a goal of biological science to identify the natural functions of some organ and structure. A biologist who says "I happen to be interested in blood circulation, therefore I see the heart's function in pumping blood" would appear rather unusual. Biologists want to discover what the function of some biological structure is, and they want their functional explanations to be made true by natural facts. Thus, Cummins' desiderata for functional analysis and those of a modern biologist appear to be different. (Weber 2005a, 191)

So, why doesn't Weber just adopt the etiological SE approach to biological functions as his objective approach? Clearly, as we shall see, Weber, like Cummins, is driven by a broader agenda related to philosophy of mind—in this case, the idea that forms of holism may be common to philosophy of mind and philosophy of biology. One might wonder if the rivalry in philosophy of mind between classic CR functionalism and Millikan's SE-function analysis of meaning has eventuated in what amounts to a proxy war in philosophy of biology!

Weber is clearly opposed to any etiological or historical loading of "function," and he does offer a couple of arguments against the SE view as a univocal approach:

> First, it will not admit anything as a function that has just arisen anew (for example, by spontaneous mutation) without having experienced the influence of natural selection yet. Second, biologists sometimes attribute functions without knowing the evolutionary past of some part or structure. (Weber 2005a, 191)

These are weak reasons to give up the use of natural selection in determining biological function. In the limit case of the initial occurrence of a novel spontaneous mutation that accidentally confers a new capacity, it is usual to say that the novel effect is not its natural function but a fortuitous benefit. People who, say, develop a chance mutation that makes them impervious to HIV infection are not said to possess a mutation that has the natural function of protecting them from HIV. Rather, by happy accident, they have a mutation that protects them from HIV and "functions to" (if you will) protect them.

As to the point that biologists often attribute functions without knowing the evolutionary past (a point made also by Murphy and by many other critics), of course they do. Even when there are attributions of function, they are generally based on inferences about biological design from circumstantial evidence rather than knowledge of evolutionary history. Evolution is an essentialist *theory* of natural functions, not the meaning

of "function" (see my reply to Lemoine in this volume). One can do plenty of biological and psychiatric science without knowing evolutionary histories (see my reply to Kincaid in this volume) or even without knowing about evolution. The notion of a natural or biological function was understood long before Darwin and is understood today by those who reject or are ignorant of Darwinian theory. Aristotle did not need to know about evolution to judge that a function of the eyes is to enable one to see, of the hands to grasp, of fear to evade danger, or that teleology was somehow involved in acorns regularly turning into oak trees. Weber's argument that people judge functions without knowing evolutionary history, and therefore functions cannot be naturally selected effects, confuses sense and reference and is as fallacious as arguing people often judge liquids to be water without knowing chemistry, and therefore water cannot be H_2O. One does not have to know that water is H_2O to start a useful science of hydrology, and one does not have to know that biological functions are naturally selected effects to recognize natural functions and even to inaugurate a useful profession of medicine that tries to help people when something goes harmfully wrong with biologically designed functioning.

Note that in constructing his objective version of CR functions to match the standard objective usage he describes, Weber is not claiming that, if he can make his idea about objective CR functions work, then there is no need for some other objective notion such as SE functions. Like most other defenders of CR functions, Weber allows that "function" may be polysemous and that there is room for more than one meaning in different contexts. For example, he says that "to make sense of scientific practice it is necessary to give an account (or several accounts, should there be different concepts of function used in biology) that picks out those things as functions that biologists ascribe functions to" (191) and remarks that it is "questionable" if there is one standard usage of "function" in biology (191). This is the same openness to other than CR functions that we saw in many of the authors cited by Murphy.

I now turn to Weber's proposed solution to Murphy's question of how CR functions can finesse interest relativity and successfully mimic a naturalist account of biological functions without reference to natural selection or other etiological historical criteria. Weber attempts to do this by developing a position within functional biology that is analogous to coherentist views of the justifiability of beliefs in epistemology, where it is the internal coherence of a system of beliefs that allows one to judge it as "true." Analogously, Weber adds a "coherence" requirement to the CR account's function attributions that is intended to yield objective judgments without any reference to external criteria such as evolutionary history: "this account can be supplemented with a coherence condition in order to avoid a certain kind of relativity to the investigator's interest" (2005a, 190).

Weber argues that although there are an indefinite number of causal capacities of a mechanism that might be of interest, only certain of the mechanism's causal capacities are part of a larger interacting coherent system of capacities of various mechanisms

that constitute the system's incredibly complex interacting system of functions, and it is those coherence-manifesting effects and only those that constitute a mechanism's biological functions: "the parts of a system have some of their characteristic properties—namely, their biological functions—only because they form a coherent system with other components that have biological functions" (200). This offers what Weber initially hopes is an objective and naturalistic way of establishing which of a mechanism's many effects on which of the organism's many capacities are the mechanism's natural functions.

The coherence analogy suggests to Weber that he can analyze functions by means of coherence: "X's function in system S is φ exactly if X's capacity to φ coheres with other capacities belonging to (parts of) S. The concept of coherence as understood here designates a complex relation between a large number of capacities. The basic relation on which this coherence relation is based consists in a capacity's contribution to another capacity. The exemplary case is the heart's contribution to the circulatory system's capacity to transport solutes and cells through the body" (192–193). Weber further explains the notion of coherence and illustrates it as follows:

> Let us say that a system of capacities is coherent if it contains a sufficiently complex net of such contributory relations between the various capacities, such that many capacities contribute to other capacities that contribute themselves to other capacities and so forth. … Biological organisms contain an elaborate network of capacities that contribute to other capacities. Here is just a small section through such a network: The function of certain ion channels in nervous membranes is to regulate ion permeability because this capacity is part of an account of the nervous membrane's capacity to fire action potentials. But the nervous membrane's capacity to fire action potentials is part of an account of the nervous system's capacity to process information. Therefore, it is a function of nervous membranes to fire action potentials. Furthermore, the nervous system's capacity to process information is part of an analytic account of the organism's capacity to locate food and sexual partners. Therefore, it is a function of the nervous system to process information. The organism's capacity to locate food is part of an analytic account of its capacity to ingest energy-rich compounds and nutrients, which are part of analytic accounts of the liver's capacity to synthesize purines and pyriminides and of the muscles' capacity to transform chemical energy into motion. (193–194)

Weber observes,

> It is obvious that biologists could tell many endless stories like this one. Any organism of some complexity will reveal zillions of such explanatory relations; this is what it means to possess a functional organization (and perhaps, to be an organism). What I am suggesting here is that, if there is a unique way of laying such a coherent functional organization over an organism it is the place of a given capacity in such a coherent system that underwrites this capacity's status as a function, and not its selection history nor the investigator's interests. (193)

However, Weber is immediately forced to supplement the coherence account. The analogy to belief systems offers the first clue of trouble. To take the analogy further, the standard objection to the coherence theory of truth is that truth is objective, whereas

there can be different mutually contradictory equivalently coherent systems of beliefs. Optimizing coherence of beliefs is thus not necessarily the same as optimizing the truth of beliefs. For example, for all we know, certain all-encompassing religious dogmatic belief systems or even some delusional systems or pathological belief systems may contain many false beliefs and yet may be as coherent as belief systems that contain many more truths. Against Weber's coherence theory of functions, one might raise a similar objection—in effect, an analog of Bertrand Russell's (1907) classic objection to the coherence theory of truth that there are equally coherent systems of belief that go along with a proposition and its negation, so coherence per se cannot constitute truth. Analogously, given a disease that does not reduce life expectancy, a causal relation and its negation may be part of equally coherent and self-maintaining systems of CR functions. The claim that health is more coherent than illness is not immediately persuasive given that there is no reason to think that maintaining a chronic disease state in a living human being involves any less causal pathways to organism features than does maintaining a healthy state. As Weber emphasizes, organisms are complex systems of interacting natural functions that have a remarkable coherence, and, interacting with parts of that coherent system, there can be equally remarkable coherent subsystems of dysfunctions induced during some disease processes that, for example, allow cancer to spread despite all the genetic safeguards, allow infections to persist despite the immune system, and allow parasites to coexist within us and adjust our internal environments to their own benefit. At the very least, the level of coherence in disease versus health is an empirical question, so a coherence view implies that it is an empirical question whether each disease is in fact a natural function, a position impossible to defend.

A second strike against the coherence theory is that it does not explain the continuity in the attribution of biological function over historical time. Weber's examples of coherence are all examples of the amazingly complex workings of the human body that we have come to understand relatively recently, as in the above quoted example. However, biologists identified objective functions long before they had an inkling of how complicated we are at the many levels we now understand. In this regard, it is worth comparing two explanations for why we judge the heart's function to be pumping the blood. Here is Weber's explanation, in which he presents a more convoluted version of Hempel's (1965) classic dilemma of why pumping the blood is the function of the heart, whereas making a sound is not:

> The heart has the capacity to pump blood, which contributes to the circulatory system's capacity to deliver oxygen and nutrients to all body cells. But the circulatory system does many other things: For example, it delivers signaling molecules such as hormones and removes metabolic waste from the cells for chemical decomposition in the liver or dialytic removal in the kidneys. It also carries platelets (for repair), antibodies and immune cells such as B- and T-lymphocytes through the body. For simplicity, let us treat these various activities of the circulatory system as one capacity, the transport capacity of the circulatory system. The question now is whether

> biologists have chosen this capacity just so, because they happen to be interested in transport. This seems not right. Intuition prompts us to say that the transport capacity is the salient capacity of the circulatory system. The circulatory system also generates heat and carbon dioxide, uses up energy-rich compounds, makes noises, forms blood clots and hence causes disease and death, but these capacities are not salient.
>
> But why is the transport capacity salient? An obvious answer is that the transport capacity is the circulatory system's function, while generating heat and carbon dioxide, using up energy-rich compounds, making noises and forming blood clots are not. … This raises the obvious question of what underwrites the functional status of the circulatory system's transport capacity. Perhaps it is the fact that the transport capacity contributes to a variety of other capacities that are also functions: cell respiration, immune defense, catabolic waste removal, metabolic coordination, sexual differentiation, and so on.

Now, what is salient about this passage is that we know that Sir William Harvey (1628/1993) identified pumping the blood as the function of the heart without knowing any of those other specifics. He knew nothing about coherence of the heart's pumping with the many other processes mentioned by Weber. So, why did he conclude, correctly, that pumping the blood (i.e., blood transport) is the function of the heart? No doubt, mostly biologists are trying to figure out how things work without worrying about strong function statements. Indeed, as Tinbergen (1963) famously noted, a mechanistic understanding of how things work is integral to an overall evolutionary explanation (I agree; see my reply to Gerrans in this volume). Indeed, one cannot really see the ways that functional attributions and explanations are needed until one has a grasp of the descriptive facts that allow one to conclude that there is a design-like quality that makes the existence of natural functions likely. For example, Harvey's discovery that the function of the heart is to circulate the blood is considered by many to be the greatest single medical discovery of all time (Friedland 2009) and took place more than 200 years before Darwin's discoveries. Yet it is clear that in attributing a primary function of pumping blood to the heart, Harvey was influenced by strong intuitions about biological design, although he did not yet know what explained biological design:

> From the symmetry and magnitude of the ventricles of the heart and of the vessels entering and leaving (since Nature, who does nothing in vain, would not have needlessly given these vessels such relatively large size), from the skilfull and careful craftsmanship of the valves and fibres and the rest of the fabric of the heart, and … how great the amount of transmitted blood … I began privately to think that it might rather have a certain movement, as it were, in a circle. … It must therefore be concluded that the blood in the animal body moves around in a circle continuously and that the action or function of the heart is to accomplish this by pumping. This is the only reason for the motion and beat of the heart. (Harvey 1628/1993, as quoted in Ribatti 2009, 1–2)

Harvey focuses on certain features of form and causal power, as CR theorists insist, but, contrary to CR theory, he is not primarily interested in sheer causal or capacity

attributions except as a path to understanding design. It is not the causal features and capacities for blood movement of the ventricles per se that capture his attention, but, in light of relatively simple and primitive observations of structure combined with the assumption of biological design ("Nature, who does nothing in vain, would not have needlessly given these vessels such relatively large size"), and from detailed observations of the nature of design-like features ("from the skilfull and careful craftsmanship of the valves and fibres and the rest of the fabric of the heart"; Harvey had experimented with the way the valves force blood in one direction), he concludes, first, mechanically and contrary to the standard thought of his day, that the blood "might rather have a certain movement, as it were, in a circle" before ultimately concluding with the stronger statement that "the action or function of the heart is to accomplish this by pumping," using "reason" as a term connoting the purpose for which the heart is there ("This is the only reason for the motion and beat of the heart"). In light of endless such historical examples (e.g., Aristotle had no understanding of the myriad causal linkages necessary to get from an acorn to an oak tree but nevertheless understood the production of the oak tree as the final cause, or function, of the acorn), it would appear that Weber's appeal to complex coherence relations is an ad hoc appeal to things we know now long after our notion of biological function became a regular part of biological theorizing.

To the degree that coherence seems likely in biologically designed organisms, Weber seems to have gotten the relationship between design and coherence backward. The history of the study of the heart illustrates that in our ignorance of mechanisms and their natural-selective history, coherence is generally implicitly assumed based on an assumed teleology, not the other way around. For example, we don't really know yet why we need to sleep (or at least we didn't until recently; there are some exciting developments in this area). Thus, we don't really know the degree to which sleep is all that coherent with the rest of our functions. Yet from circumstantial evidence, we assume that we are biologically designed to sleep and so it is normal, and we presume that likely it links to many other functions. This is despite the fact that sleep takes away a third of our lives, leaving us during that time largely unaware of our environment, functionally impaired, periodically hallucinating, and partially paralyzed. Yet sleep isn't considered a disorder. This would remain true even if we were to discover that sleep is not very linked to other functions but a unique one-off naturally selected adaptation that could be suppressed without harm (e.g., as is usually the case with our biologically designed capacity to develop a fever in response to an infection). A lack of high coherence of the kind Weber describes would not make our need for sleep any less normal given that it is biologically designed.

Furthermore, high degrees of coherence can be achieved by socially constructed functions that are not natural or biological functions. Indeed, cultures often support their ideologies by claiming that certain functions that have become "second nature" within that culture are natural to human beings, when in fact those functions have been locally

constructed by the culture. Because such ideologies can serve as rationales for oppression of deviant individuals and misclassification of such individuals as disordered, drawing the distinction between biological functions and other forms of socially cultivated functioning is a major goal of an analysis of the concept of mental disorder. The coherence view fails to accomplish this goal because socially cultivated functioning can become integrated in overall functioning in a coherent fashion. For example, Kingma (2013) observes that a social invention such as reading can be so integrated with other functions that it can appear naturally selected until we recall that it is an invention.

The coherence account is also subject to counterexamples of a kind that might be called "function danglers." Surely there can be functions of specific single organismic features that have relatively isolated causal chains leading to their final result and so do not interact with many other features. Worms apparently have light-sensing cells on their tails, and thus one imagines perhaps that they feel a sense of satisfaction when finally fully underground, but that function of the light-sensitive cells may interact with nothing else on its way to reinforcing the safety of being underground. So, its level of coherence would be quite low. One can imagine even more extreme cases, such as external markings that have a function in terms of potential predators but do not interact at all with other features of the organism. Such "danglers" possess biological functions just as fully as the highly coherent ones described by Weber, yet they are not significantly coherent.

The Self-Reproduction Criterion

Weber recognizes that uniqueness of function is not assured by his coherence criterion, and he tries to obtain uniqueness by adding another criterion, "self-reproduction," by which he means the organism's continued existence over time. This criterion, borrowed from McLaughlin (2001), is intended to further whittle down the effects that define the privileged coherent system that equates to objective biological functions:

> The crucial question is obviously whether there is a *unique* coherent system of capacities. Doubts are in order; it is quite conceivable that there are many ways of knitting various causal dispositions of the parts of an organism into a coherent system in the manner just outlined. However, what seems less likely is that there are several systems that are *explanatorily equivalent*. It is possible that, for any type of organism, there exists exactly one coherent system of capacities that best explains how the organism can self-reproduce. By "self-reproduction" I mean not procreation, but the organism's capacity to maintain its form or identity for a certain appropriate duration (see McLaughlin 2001). This appears to be the most universal property in biology (note that not all organisms procreate!), and it is certainly the property that biologists ultimately want to understand. For these reasons, it is appropriate to take self-reproduction as the capacity that a system of functions must explain. (194)

In sum: "I conclude that a biological (role) function is a capacity that either contributes to a higher-level system capacity that is itself a role function or contributes directly

to an organism's self-reproduction" (2005b, 40). Interestingly, the view that inspired Weber's use of self-reproduction is an etiological view, which he rejects, so he must amend it. Weber explains,

> A different kind of etiological account of functions has recently been developed by Peter McLaughlin (2001). On McLaughlin's view, … A system can only have functions if it is capable of self-reproduction. By this, McLaughlin means the maintenance of an organism's form over time, which is not to be confused with … procreation … by constantly regenerating their parts. My body does not contain the same carbon atoms as it did 20 years ago, but the individual organism that I am sustains metabolic activities continually replacing all the atoms and molecules that make up my body. (2005b, 37)

Why is this an etiological account despite no involvement of natural selection? The basic idea is simply that if a mechanism in an organism contributes to the survival (or "self-reproduction") of the organism, then because the organism survives, the mechanism also survives within it, and thus there is a sense in which the mechanism's effect explains its own continued existence: "McLaughlin argues that there is a sense in which tokens of function bearers, by virtue of what they do, indirectly cause their own continuing presence in an individual organism, namely by contributing to the maintenance of the whole system (token). Thus, McLaughlin's account can be classified as etiological" (2005b, 37). Weber firmly abandons any etiological element of the self-reproduction account, distinguishing his use of self-reproduction from that of McLaughlin who inspired it: "It is important to appreciate that, on this account of functional explanation, the ascription of a function to a biological entity implies nothing about why tokens of this entity are present in the systems that are given a functional analysis. … McLaughlin's version of the etiological function concept … does not correctly capture the use of the term 'function' … because, as I have argued, functional analysis does not at all attempt to explain why tokens of function bearers are present in a biological system" (40). Thus, it is McLaughlin's account of functions as CR contributors to self-reproduction without the explanatory element.

One might be dubious from the start that one can define the criterion of self-reproduction without a prior understanding of health and thus of function. What constitutes the acceptable perpetuation of the organism depends on the proper functioning of the organism's parts, or else any horrific disordered state that maintains life would qualify all the parts as functioning properly. This misunderstanding that one can analyze the functions of parts in terms of organismic maintenance goes all the way back to Hempel's (1965) classic paper on functions in which Hempel at one point proposed that an effect of an organ is a function if it "ensures the satisfaction of certain conditions … which are necessary for the proper working of the organism" (305). As Cummins (1975) observes, this is circular as an analysis of "function" because "it seems clear that for something to be in working order is just for it to be capable of performing its functions, and for it to be in adequate or effective or proper working order is just for it to be

capable of performing its functions adequately or effectively or properly" (753). "Self-reproduction" does not even seem to require self-reproduction in a properly working way and so is even more subject to Cummins's objection than is Hempel's claim.

These broader concerns aside, Weber's coherence account is not saved from ample counterexamples by the arbitrary addition of the self-reproduction filter to identify objective natural functions. One immediate set of counterexamples consists of those reproductive organs and features concerned solely with reproduction that do not contribute to self-reproduction and yet have functions. A second set of counterexamples involves those many serious illnesses that in our modern environment do not influence life expectancy and so on the basis of self-reproduction alone cannot be classified as disorders.

A third set of counterexamples, noted above, consists of the many ways that humans culturally exploit their natural mechanisms so that cultural practices become a seamless part of a coherent and interacting set of capacities supporting self-reproduction. In such cases, a capacity is part of the coherent system supporting self-reproduction but is not considered a biological function, yet Weber's account would classify it as a function. For example, the ability to learn to read is not a biological function of any human mechanism, yet it is difficult to imagine a capacity that enters into our lives in modern literate societies more coherently as part of a larger interacting system of capacities that keep us alive.

There are also mechanisms with biologically designed capacities that possess functions but are not currently contributing to self-reproduction due to changes in the environment. For example, the preference for fatty and sweet foods is the function of certain taste and hunger mechanisms, yet in our current environment, we spend our lives trying to minimize, distance ourselves from, and otherwise control that capacity from expressing itself in behavior because with our plentiful food supply, these tastes yield lower self-reproduction via disease causation. Yet other clear counterexamples to the self-reproduction view emerge from the theory of kin selection, which implies that some naturally selected functions may go against self-reproduction and encourage individuals to sacrifice themselves to save cofamilial individuals who have a sufficient share of the same genes, thus increasing the likelihood of gene perpetuation at the cost of ending self-reproduction. One can think here of the seemingly programmed non-self-reproduction of salmon after spawning that provides nutrients to offspring or the ready sacrifice of soldier ants.

I conclude that Weber's coherence account, even with the self-reproduction epicycle added, does not save CR functions from its fatal flaws as an account of what we mean by natural function in the context of biological design. The problem here is not the details of the particular proposals. The problem is the entire strategy of starting with the anemic CR account of function and trying through various restrictions to get the result to come out equivalent to the robust SE account of functions that accurately

reflects the objective sense of biological function. Like analyzing "human being" as "featherless biped," one might get close to material equivalence, but one would still not be achieving an understanding of how we think about the concept, why it is objective, and why we take it so seriously.

Conclusion

The examination of the work of Craver and Weber that Murphy cites shows that in fact those sources fail to provide any solution to the problem of how CR functions can explain the objectivity of medical dysfunction. This means that Murphy fails to directly show or indirectly indicate how CR functions can undergird an objective sense of dysfunction and disorder in medicine and psychiatry. Murphy thus fails at his self-appointed task of showing how CR functions can explain when something has objectively gone wrong "under the hood" of the human being. The "scientific image" he creates of psychiatry based on CR functions is in fact unscientific and subject to all the abuses that the analysis of the concept of mental disorder was undertaken to address. And, the HDA remains the only scientifically grounded account that fulfills the objectivity criterion that he accepts.

References

Amundson, R., and G. V. Lauder. 1994. Function without purpose: The uses of causal role function in evolutionary biology. *Biology and Philosophy* 9: 443–469.

Bock, W. J., and G. von Wahlert. 1965. Adaptation and the form-function complex. *Evolution* 19: 269–299.

Craver, C. F. 2001. Role functions, mechanisms, and hierarchy. *Philosophy of Science* 68(1): 53–74.

Craver, C. F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press.

Cummins, R. 1975. Functional analysis. *Journal of Philosophy* 72: 741–765.

Friedland G. 2009. Discovery of the function of the heart and circulation of blood. *Cardiovascular Journal of Africa* 20(3): 160.

Harvey, W. 1628/1993. *On the Motion of the Heart and Blood in Animals*. Trans. R. Willis. Prometheus Books.

Hempel, C. G. 1965. The logic of functional analysis. In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science,* 297–330. Free Press.

Kingma, E. 2013. Naturalist accounts of mental disorder. In *The Oxford Handbook of Philosophy and Psychiatry,* K. W. M. Fulford, M. Davies, R. G. T. Gipps, G. Graham, J. Z. Sadler, G. Stanghellini, and T. Thornton (eds.), 363–384. Oxford University Press.

McLaughlin, P. 2001. *What Functions Explain: Functional Explanation and Self-Reproducing Systems*. Cambridge University Press.

Ribatti, D. 2009. William Harvey and the discovery of the circulation of the blood. *Journal of Angiogenesis Research* 1(3): 1–2.

Russell, B. 1907. On the nature of truth. *Proceedings of the Aristotelian Society* 7: 228–249.

Salmon, W. S. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.

Salmon, W. S. 1998. *Causality and Explanation*. Oxford University Press.

Tinbergen, N. 1963. On aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20: 410–433.

Wakefield, J. C. 2003a. The Chinese room argument reconsidered: Essentialism, indeterminacy, and strong AI. *Minds and Machines* 13: 285–319.

Wakefield, J. C. 2003b. Fodor on inscrutability. *Mind and Language* 18: 524–537.

Weber, M. 2005a. Holism, coherence and the dispositional concept of functions. *Annals of the History and Philosophy of Biology* 10: 189–201.

Weber, M. 2005b. *Philosophy of Experimental Biology*. Cambridge University Press.

Defining Mental Disorder

Jerome Wakefield and His Critics

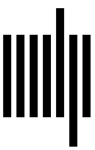
By: Harold Kincaid, Peter Zachar, Dominic Murphy, Justin Garson, Philip Gerrans, Rachel Cooper, Steeves Demazeux, Leen De Vreese, Maël Lemoine, Tim Thornton, Andreas De Block, Jonathan Sholl

Edited by: Luc Faucher, Denis Forest

Citation:

Defining Mental Disorder: Jerome Wakefield and His Critics
By: Harold Kincaid, Peter Zachar, Dominic Murphy, Justin Garson, Philip Gerrans, Rachel Cooper, Steeves Demazeux, Leen De Vreese, Maël Lemoine, Tim Thornton, Andreas De Block, Jonathan Sholl
Edited by: Luc Faucher, Denis Forest
DOI: 10.7551/mitpress/9949.001.0001
ISBN (electronic): 9780262362931
Publisher: The MIT Press
Published: 2021

The open access edition of this book was made possible by generous funding and support from Arcadia – a charitable fund of Lisbet Rausing and Peter Baldwin



The MIT Press

The open access edition of this book was made possible by generous funding from Arcadia—a charitable fund of Lisbet Rausing and Peter Baldwin.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Faucher, Luc, 1963– editor. | Forest, Denis, editor.
Title: Defining mental disorder : Jerome Wakefield and his critics / edited by Luc Faucher and Denis Forest.
Description: Cambridge, Massachusetts : The MIT Press, [2021] | Series: Philosophical psychopathology | Includes bibliographical references and index.
Identifiers: LCCN 2020016671 | ISBN 9780262045643 (hardcover)
Subjects: LCSH: Wakefield, Jerome C. | Psychiatry--Philosophy. | Mental illness--Philosophy. | Mental illness--Diagnosis. | Mental illness--Classification.
Classification: LCC RC437.5 .D434 2021 | DDC 616.89--dc23
LC record available at https://lccn.loc.gov/2020016671