



Project
MUSE[®]

Today's Research. Tomorrow's Inspiration.

Mental Illness, Human Function, and Values

Megone, Christopher.

Philosophy, Psychiatry, & Psychology, Volume 7, Number 1, March 2000, pp. 45-65 (Article)

Published by The Johns Hopkins University Press



 For additional information about this article

<http://muse.jhu.edu/journals/ppp/summary/v007/7.1megone.html>

MENTAL ILLNESS, HUMAN FUNCTION, AND VALUES

CHRISTOPHER MEGONE



ABSTRACT: The present paper constitutes a development of the position that illness, whether bodily or mental, should be analyzed as an incapacitating failure of bodily or mental capacities, respectively, to realize their functions. The paper undertakes this development by responding to two critics. It addresses first Szasz's continued claims that (1) physical illness is the paradigm concept of illness and (2) a philosophical analysis of mental illness does not shed any light on the social and legal role of the idea. Then, in reply to Wakefield, the aim is to defend the account as an interpretation of Aristotle and to argue that this Aristotelian view of mental illness is preferable to one that rests on a supposed value free account of human function. More generally the discussion points to the fact that both Wakefield and Szasz rely on a number of metaphysical assumptions about the supposedly empirical nature of medical diagnosis, about the relation between facts and values, and about mind and body (among others), which are open to challenge. In particular the paper indicates an Aristotelian approach to the fusion, in the natural world, of so-called *facts* and *values*, and the relevance of this fusion to the analysis of the concept of illness. This suggests the debate over distinct conceptions of that concept must both illumine and be illuminated by these deeper metaphysical questions.

KEYWORDS: fact, nature, physical illness, rationality, teleological explanation

I AM MOST GRATEFUL for the work of Professor Wakefield and Professor Szasz, as well as that of previous commentators, in response to my paper on Aristotle's function argument

and mental illness. The debate here emphasizes the value of the cross-fertilization made possible by a journal such as *Philosophy, Psychiatry, & Psychology*, and the need for such a journal. In the present case, my fellow contributors are both much better qualified than I in the psychiatric field, and this is of benefit to my work. However, from my perspective, both seem unduly incautious in some of their philosophical claims, which may leave scope for my contribution. Although adopting very different positions, both Wakefield and Szasz seem to rely on some doubtful general assumptions. In Szasz's case one might cite an apparent adherence to Cartesian dualism about mind and body, and assumptions about meaning and definition, and method in establishing these. In Wakefield one might note implicit views as to a theory of explanation and about the nature of essentialism. Finally both hold views about the supposedly empirical nature of medical diagnosis, about values, and about the relation between facts and values, which require scrutiny.

The aim of this paper is to respond to the charges made against my position by Szasz and Wakefield. However, in so doing, it should also be possible to illuminate, and question, the role played by assumptions of the sort mentioned above. More positively still, the paper should indicate, at least, the way in which an Aristotelian account of the natural world fuses so-called

facts and *values* (the terms are not Aristotelian) and also the way in which an analysis of the concepts of illness and health illustrates the advantages of such an account.

In this first section the charges made against my Aristotelian account will be outlined and an indication of the response will be given. The second section will examine further Szasz's claims about the concept of physical illness. Since Szasz surprisingly admits in this discussion that in his view the concept of physical illness refers to functional norms, this introduces the question of the role of functional norms in an analysis of the concept of illness. The third section provides a clarification of the Aristotelian account previously presented, responding to a range of criticisms made by Wakefield. In the fourth section, Wakefield's unusual conception of teleological explanation and his proposed alternative account of natural function will be rebutted. These third and fourth sections provide space for an indication of the fact/value relation in the Aristotelian approach. In the fifth section, in response to Szasz's charge, the importance of the philosophical analysis of the concept of mental illness for the social and legal use of that concept will be considered. Finally a way forward for developing the Aristotelian account of the concept of mental illness is briefly sketched and the joint role of philosophy and psychiatry in such a project is indicated.

What then are the claims that must be addressed? Szasz holds that the Aristotelian position put forward has not refuted his argument that the concept of mental illness lacks literal meaning and that the concept of physical illness is the paradigm concept of illness (Szasz 2000). Furthermore he continues to assert that a philosophical analysis of mental illness is irrelevant to understanding the social and legal role of the idea, which is what most requires attention (*ibid.*). Wakefield, on the other hand, claims that the interpretation given of Aristotle's function argument (which he takes to be a misinterpretation) in effect holds that judgments about mental disorder (illness) are essentially value judgments, and thus is a position "on the same side as Szasz with respect to the most important issues"

(Wakefield 2000, 42). In Wakefield's view, this undesirable state of affairs can be remedied by providing a value-free account of natural (human) functions (an account he seems to attribute to Aristotle), which can then constitute a non-evaluative component in an analysis of mental disorder.

In response to Szasz's first claim above, it may be helpful at the outset to indicate the structure of the argument previously deployed against his first point. Szasz has asserted that physical illness is the paradigm concept of illness. This seems to imply that the meaning of this concept is clear and undisputed. The first stage of the reply (1) simply followed Fulford and others in noting that the literature does not reveal a clear and well understood concept of physical illness (Fulford 1991, 80–81; 1998, 216). Given this, there is no reason to give that concept a privileged status. (2) By contrast, since both mental and physical illnesses are widely believed to exist, an account of the concept of illness revealing the two concepts to have something in common (though also, of course, differences) has considerable pre-theoretical plausibility. (3) An Aristotelian functional account adequately achieves this objective and also captures the evaluative connotations of the concept of illness. (4) Such a functional account also gives rise to an account of mental illness that can explain two phenomena, the abuse of psychiatric power and the misuse of drugs in treating mental illness, that any adequate account should explain. Therefore (5) such an account of illness is to be preferred to one based on the contentious (and supposedly purely physiological) account put forward by Szasz. If, as argued, all these stages are correct, this is a refutation of Szasz's claim that the concept of physical illness alone has literal meaning.

In his current discussion, Szasz has taken up stage 1 of the argument above by reasserting the paradigm position of the concept of physical illness, so that claim is reexamined, in some detail, below.¹ In response to Wakefield, it will be argued that he has radically misconstrued both Aristotle and my account of Aristotle. It will also be shown that a non-evaluative account of natural functions is implausible but that, in any case,

the Aristotelian account of natural function or of illness should not be construed as purely evaluative. There is a false contrast here. The falsity of this contrast may undercut Wakefield's own motivation for a non-evaluative account, which seems tied to his conception of science as a purely empirical enterprise. As will then be seen, apart from the conceptual problems with his account, it faces difficulties both in explaining the reason for a link between the concept of dysfunction and that of disorder, and in accounting for the source of judgments of harm that contributes to judgments of disorder. Taken together, these last points may suggest that Wakefield's own position is closer to Szasz's than he would like.

THE CONCEPT OF PHYSICAL ILLNESS

In response to the Aristotelian claim that the concept of mental illness has as literal a meaning as the concept of physical illness, Szasz reiterates his view that physical disease is the basic notion, clear and undisputed. In defense of this view, he cites the definition of disease given in the *Oxford English Dictionary*, two pathology textbooks that agree that "the basis of all disease is injury to the smallest living unit of the body, namely, the cell," and the fact that textbooks of pathology do not cite clinical depression or schizophrenia as diseases. In summary, he asserts that the core medical conception of disease is as a bodily abnormality (Szasz 2000, 4).² Given this assertion, Szasz's next move is as follows: "Extending the criterion of disease from malfunctions of the body to malfunctions of the human mind introduces a fatal infection into the materialist medical definition of disease. The mind is not a material object; hence it can be diseased only in a metaphorical sense" (*ibid.*). Thus his traditional view is reasserted.

However, there are several points to make in reply. First of all, Szasz's second move transparently depends on an unstated premise, some form of Cartesian dualism about the mind. Thus any non-Cartesian need not accept this claim. Certainly an Aristotelian account of the mind (which is not reductionist, incidentally) rejects outright

this Cartesian framework, and the conceptions of physical substance and mental substance underlying it. The Aristotelian account denies that the criterion for bodily illness is purely physical (or, for that matter, purely psychological) in that it appeals to norms that are not purely physiological. So the Aristotelian account of illness does not extend the application of a criterion from the physical to the mental. It simply applies the same criterion (which is not reducible to physiological terms) to both types of illness.

But consider now Szasz's initial claim. First of all, Szasz simply ignores the fact observed by Fulford that those who have attempted to define physical illness have not been able to provide a satisfactory definition (Fulford 1991, 80–81; Boorse 1975; Kendell 1975). Thus he avoids the tricky question of how this fact is to be reconciled with his evidence. Secondly, he does not ask himself whether any of the sources he cites are purporting to reveal the literal meaning of the concept of physical illness. Both these points raise the issue of what is being sought in a search for the literal meaning of a term and how one gets at it. Thus Fulford's point is that a range of definitions of *physical illness* have been put forward, several inconsistent with that offered in Szasz's pathology textbooks. Given this, it is not clear why the latter should be given authority.

Thus, it might be queried whether Szasz's sources are the right place, or the only place, to look for the literal meaning of a term. In the case of a dictionary or pathology textbook, what is being reported could well reflect only current usage (or medical usage), which may be confused, or at least not concerned to spell out the term's literal meaning. If so, one might suppose, in technical terms, that these sources are only giving the *nominal definition* of the term, an account of how the word is currently used, which could well be incomplete or even contain conflicts. In contrast one might argue that the literal meaning is given by the term's *real definition*, and this should be an account that fully articulates the terms' meaning, explaining ordinary usage so far as is possible and accounting for any incompleteness or conflicts there. Thus Szasz needs to be more reflective on philosophical

method, more cautious as to what's involved in giving the literal meaning of a term.

This problem with method leads to a third more crucial point, for it transpires that Szasz's own proposed literal meaning is indeed incompletely spelled out and in a way that is problematic for his position. On the basis of the pathology textbooks, Szasz claims that the core medical conception of physical illness is that of bodily abnormality. This account of the concept is compressed because it does not identify the source of the norms by reference to which abnormalities are to be picked out. What makes a particular condition of cells an abnormality? Such a literal meaning is therefore incomplete and crucially unclear. As a result, even if we allow Szasz to rely in this way on his pathology textbooks, the meaning of *disease* given there cannot be assumed to be irreducibly materialist (as he assumes in making his second claim noted above).

This point can be developed by putting it together with further remarks made by Szasz. He claims that I attributed to him the absurd view that the meaning of the concept of mental illness could be given "*without incorporating any evaluative term*" (Szasz's italics). By contrast he now states "that the concept of disease contains an evaluative element is self-evident. . . . The crucial difference between lesion *qua* bodily disease and behavior *qua* mental disease is not that one is a value-free biological fact and the other a value-laden social construct. Both are value-laden social constructs. . . . The crucial difference between bodily disease and mental disease is that what counts as a somatic pathology is based on a judgment of how the *body ought to function*, whereas what counts as psychopathology is based on a judgment of how the *person ought to function*" (Szasz 2000, 9).

This statement requires careful consideration before we can return to the purported definition *bodily abnormality*. The first point is that, in my original paper, the reason for attributing to Szasz the claim that the meaning of physical illness can be given without incorporating any evaluative terms was his assertion that "what (physical) health is can be stated in physiological and ana-

tomical terms" (Szasz 1960, 114, 116; Macklin 1973). Szasz seems now unaware that the claim that a criterion of bodily disease appeals to a norm of bodily functioning is incompatible with this earlier claim. What can be stated in physiological and anatomical terms is the physiological/anatomical condition of a being. This materialistic language does not have the resources for the language of norms, which must state whether that physiological condition is enabling the body to achieve some goal or purpose, some good. For reasons given below, it is very welcome that Szasz accepts the importance of norms of functioning in the definition of illness. However, it is important to see that this acceptance now reveals Szasz himself to be unclear as to the definition of physical illness. (Note also that the above statement about norms appears to conflict with his appeal to "*an objectively identifiable biological norm*" in a previous reply (Szasz 1998, 204).

The second point on this issue is that Szasz's remarks presuppose that the norms that govern how the body ought to function are distinct from the norms that govern how the person ought to function. If he accepts that the human body is a component of the human person, he is thus presupposing that one can determine the function of a part independently of determining the function of a whole. This is a highly contentious claim, for which he gives no defense. In fact it is extremely implausible that norms for the function of parts are distinct from norms for the functioning of the whole, as a consideration of artifacts will show. This will be discussed further below.

Recalling Szasz's earlier claim that the literal meaning of physical illness is bodily abnormality, two points are now clear. The most basic of these is that Szasz's preferred account of the literal meaning of bodily illness is incomplete and unclear. It has to be spelled out in relation to norms, and that in turn raises the question of the specification of those norms. So far as this is concerned, Szasz has left it unclear, in the case of physical illness, how it is to be determined in what way the body ought to function. However, in so doing, he has left open the possibility that the ultimate determinant of the norm of bodily

function is the same as the determinant of the norm of functioning of the whole person, since in general the function of parts will depend on the function of the whole. This would of course be compatible with the Aristotelian picture in which the evaluative norms for both physical and mental illness are the same. The second point is that in conceding the relevance of norms to the definition of both physical and mental illness, Szasz has thereby conceded that the literal meaning of physical disease, the body failing to function as it ought to function, *cannot* be given in purely materialistic language since such language does not include the notion of goals/purposes/goods that must be used to characterize norms.

So far as Szasz is concerned then, the result of the discussion above is a vindication of the first claim of my original argument. There is no reason to treat the concept of physical illness as the paradigm concept of illness since it is neither clear nor undisputed. Szasz's own favored account is not complete since the norms of bodily functioning that are part of the definition have not been specified. Furthermore the account is not clear because it now conflicts with some of Szasz's previous own claims, since the relevance of norms to the definition prevents it being specified in purely material language. (Both these points leave open the possibility that the norms specified in the meaning of physical illness are, at some level, the same as those for mental illness.)

Szasz's admission of the importance of evaluative norms of functioning for the definition of illness, whether bodily or mental, constitutes a useful introduction to the discussion of Wakefield, since Wakefield's position admits the relevance of the concept of natural functioning to an analysis of the concept of mental illness but denies that the notion of natural function is itself to be understood normatively, that is, evaluatively (thus conflicting now with both Szasz and myself).

NATURAL FUNCTIONS

The position presented in my original paper was that illness, whether bodily or mental, should be analyzed as an incapacitating failure of bodily

or mental capacities, respectively, to realize their functions. It was suggested that the notion of function here should be understood as Aristotle had understood it (incorporating, in modern terms, an evaluative component): Human beings have a function in the sense that there are goals or purposes that good human beings will realize (actualize). The Aristotelian claim is that the ultimate goal for a good human being as a whole is to live the life of a fully rational animal. The function of the bodily and mental parts of a human is to operate in ways that contribute instrumentally or constitutively to the realization of this goal. Thus a bodily ailment such as a lung condition is, at root, an incapacitating failure of the lung to function in that way whereby it contributes optimally to the life of the fully rational animal; a mental illness such as depression will involve an incapacitating failure of mental powers to be realized (actualized), thus preventing the individual reasoning optimally.

Wakefield's argument against this position has several strands. He claims that it provides an incorrect analysis of (dys)function. He also claims that it rests on a misinterpretation of Aristotle's function argument. In place of these supposed errors, Wakefield offers his own analysis of the notion of function, and of its role in the concept of mental illness, and he argues that such an account can be based on the Aristotelian function argument, properly interpreted. According to Wakefield, the notion of a human natural function should be understood non-evaluatively, though still in terms of those aspects of human behavior that can be teleologically explained. Thus the appeal to teleological explanation here purportedly requires no appeal to values. This requires a novel account of teleological explanation.

Before replying to these arguments, I will clarify my account of Aristotle's function argument, since it has been misconstrued by Wakefield. I argue that the notion of function is related to a conception of teleological explanation that entails an evaluative or normative framework. To explain how Aristotle's preferred conception of function is identified in the human case, I will indicate the way in which the Aristotelian account fuses facts

and values. I will then argue that an alternative account of teleological explanation is not needed and is not coherent.

A version of Aristotle's function argument is presented in *Nicomachean Ethics* I.7, though a further version is also given in *Eudemian Ethics* II.1, and relevant remarks are made in *Politics* I.1–3. Wakefield makes a number of objections to the interpretation previously given of the *NE* argument. (1) He holds that "Megone's interpretation turns Aristotle's claim into a triviality" (2000, 24) because the interpretation results only in claims that are true in a trivial sense. (2) The argument as I interpreted it tells the modern reader nothing about why it makes sense for reason to be the function of a human being (Wakefield 2000). (3) The account given does not adequately explain the references in the argument to the functions of flautists and carpenters (Wakefield 2000; Hobbs 1998, 211). (4) The proposed analysis of human function collapses the notion into that of the human good and thus makes the investigation of the human good in terms of *function* unhelpfully circular (*ibid.*).

At least three of these claims turn on misinterpretations by Wakefield. To see this, and to review the interpretation, it is necessary first to recall the bare bones of the argument in *NE* I.7. Aristotle argues first that there is a quite general connection between the function of a kind and the good of that kind. To support this point, he gives the examples of musicians and sculptors whose good, *qua* musicians and sculptors, is tied to the function of each skill. Given this general claim, if humans *qua* humans have a function, that will determine their good too. Aristotle then cites two considerations in favor of the view that humans as such have a function. First he suggests that if carpenters and tanners have a function, then humans as such should do also; and second, he notes that since eyes and feet have functions, the human being as a whole must do so (*NE* I.7, 1097b31ff.). His third point is that whatever the human function is, it must be distinct from that of plants and animals. And finally he concludes that the human function is realized in "an active life of the element that has a rational principle" (*NE* I.7, 1098a5ff.). It is the life of a fully rational animal.

Clearly this argument turns crucially on the claim that humans as such have a function but also on Aristotle's view of what that function is. On my interpretation, in order to understand the crucial first claim, we need to attend to Aristotle's views on natural kinds in general, which can then be applied to the case of humans (Megone 1998a, 192–94). However, according to Wakefield, in the presentation of this interpretation, function is defined in a way that renders the whole argument trivially uninteresting.

Wakefield's argument for this claim is curious. He seems to think that my interpretation rests on starting from a definition of the function of a part and then applying that definition to derive the function of the whole. Such an account would in fact run contrary to an Aristotelian principle indirectly referred to in my original interpretation, namely, that parts have functions only in so far as wholes have functions; that is, the function of the whole is the prior notion. But it is also a mistake to think the notion of function is defined primarily with reference to either part or whole. Wakefield gives the basis for his position as follows.

(1) Megone construes Aristotle as arguing "that if parts of a human, such as the eye or the foot, have a function then a human being as a whole must do so" (Megone 1998a, 192). That is, he suggests that it is conceptually necessary for the whole organism to have a function if a part does. . . . (2) Recall that Megone defines a part's *function* as whatever the part contributes to the life cycle changes of a good member of the species. Thus, he thinks that "[f]unctional explanation only makes sense in the light of the function of the whole" (195) because each function by definition must contribute to the organism's overall functioning. (3) Indeed, if one mechanically applies Megone's definition of *function* to the whole organism rather than to a part, one finds that the function of the organism is whatever the organism contributes to the characteristic life of the organism, which is everything! (Wakefield 2000, 23).

As indicated already, there are two basic problems here: One concerning the definition of function, the other concerning the relation of parts and wholes. Take attribution 1 first. At no point was it suggested that it is conceptually necessary for the whole organism to have a function if a

part does. In fact the idea that a claim about the whole organism is conceptually dependent on a claim about the parts gets things just the wrong way round and gives a contentious account of the mode of argument. On the contrary, Aristotle's suggestion (not an explicit argument) (*NE.I.7*, 1097b32–33), that if the eye or foot has a function, then the whole human must do so, is *based* on a suppressed general principle that parts can only have functions if they belong to wholes that have a function. According to this principle, the functions of wholes are prior to the functions of parts. Given such a general principle and the fact that we pre-theoretically believe that eyes, for example, do have functions, namely, to see, we are driven to accepting that human beings as such have a function.

The basis for the general principle on which the argument depends was not indicated in the previous paper, but it is certainly asserted by Aristotle and looks like a metaphysical claim about the nature of parts and wholes, not a conceptual claim.

We may now proceed to add that the city is prior in the order of nature to the family and the individual. The reason for this is that the whole is necessarily prior to the part. If the whole body is destroyed, there will not be a foot or a hand, except in the ambiguous sense in which one uses the same word to indicate a different thing, as when one speaks of a 'hand' made of stone; for a hand when destroyed will be no better than a stone 'hand.' All things derive their essential character from their function and their capacity, and it follows that if they are no longer fit to discharge their function [*because* no longer part of a functioning whole], we ought not to say that they are still the same things, but only that, by an ambiguity, they still have the same names. (*Politics* I.2, 1253a18–25)

This general principle certainly has plausibility if we consider an artifact such as a clock. The function of any part of the clock is derivative from the function the clock as a whole is supposed to perform. The goal of each part is to contribute instrumentally or constitutively to the clock's purpose of telling the time. The function of the whole is prior to that of the part, not derived from the function of a part, as Wakefield claims.

The second claim (2) makes an incorrect attribution concerning the definition of *function* as a

part that contributes to the life cycle changes of a good member of the species. Elsewhere Wakefield inconsistently attributes the (different) claim that "human function is itself defined as whatever it is about human nature that leads to a good human life" (Wakefield 2000, 29). In fact I do not stipulate either of these definitions since I do not define function by reference to parts. These claims may turn out to be true given the definition of function that is given and an investigation of human nature, but they are not true by definition. I shall turn to my actual definition in due course.

It is not clear that Wakefield's third attribution (3) follows, even were he not incorrect in the first two attributions. At least two errors seem clear. First, he seems to think he can infer something about the content of the function of the whole organism from a definition of the function of a part. This error does depend on the misattributions. First I do not begin by defining function by reference to parts, so that cannot be the basis of an inference to the function of the whole. Second, as has now been made clear, the attributed inference is exactly the opposite of the form of inference permitted by the principle stating the priority of wholes over parts given above. The second error is that he seems to think that the function of a whole is going to be the same kind of thing as the function of a part. But it is not at all clear this is the case. The function of a whole is to do something, a clock tells the time; the function of a part is to enable the whole to do something, the hand of a clock rotates at a certain speed in order that it contributes optimally to the clock's telling the time. The function of the whole determines what the parts are for, while the functions of parts contribute to the achievement of that overall function (Charlton 1992, 120–21).

As a result, Wakefield's concluding assertion, the "definition [of the function of a part] trivially implies that every organism has a function" (2000, 23) is opposite to what my interpretation of the function argument holds, which is that the function of the whole human determines the function of parts of humans. Thus when Aristotle claims that the function of a human is a fully

rational life, he is not talking about the function of a part, but the function of the whole that determines the function of parts. If my interpretation renders the function argument as wholly nontrivial, what does it mean to claim that a human being has a function? In answering this, it will be possible to reiterate what exactly I do take the Aristotelian definition of function to be.

The basic move in my interpretation was to claim that a human being has a function in the sense that any member of a natural kind has a function. Aristotle's argument locates human beings within his general metaphysical account of natural kinds and both that move and that picture of the natural world are wholly nontrivial. Thus in the previous paper, I asserted that in Aristotle's view, members of natural kinds have a subset of potentialities that constitute their essence or nature. For example, an acorn, or any member of the oak species, should be thought of as, essentially, a set of powers, ways of changing (acting on other things or being acted upon), where this is a subset of all the powers the acorn, or oak, has. Aristotle clarifies what powers are in this subset by noting that these essential properties (essential powers) are those whose "realization [actualization] is open to teleological explanation, and it is in this sense that the natural kind has a function" (Megone 1998a, 193). In other words, to claim that a natural kind has a function is to claim that there are aspects of its behavior that reveal its essence, and those aspects are open to teleological explanation. Given the derivative nature of the functions of parts, parts of a natural kind have functions insofar as at least some of the behavior of parts can be explained teleologically as contributing to the function of the whole. Thus the favored account of the function of a thing is that the function is *that aspect of a thing's behavior (whether the thing be a whole or a part) that is open to teleological explanation.*

Comparison with an artifact may help again. An artifact, a bed or a knife, say, has a function in the sense that there is some goal or purpose that the designer (or sometimes the user) has given it. This goal is the designer's aim in making it and aspects of a knife, or its behavior, or the behavior of its parts, can be explained teleologically by citing this goal. For example, a knife's

being sharp can be explained teleologically in terms of the goal of cutting. It is just this feature, the possibility of explaining a thing's behavior in terms of a goal, that Aristotle transfers to the case of natural kinds to justify the use of the term function there. He admits that appeal cannot be made to the goal of a designer as is made in the case of artifacts (*Physics* II.8, 199b27–30), but he argues that nonetheless goal-based explanation is applicable to nature, and in that sense members of natural kinds have functions. Thus the sense in which both artifacts and members of natural kinds (and the parts of each) have functions is just that some aspects of the behavior of (both parts and wholes of) each is teleologically explicable. This makes clear that on this interpretation, with this account of function, the Aristotelian argument that humans have a function is not at all trivial, since Aristotle was well aware of the need to argue carefully for the claim that some of the behavior of natural kinds requires teleological explanation. Just as nowadays, there were in Aristotle's time materialists who denied this claim and Aristotle addresses a number of arguments to their position (*Physics* II.8–9).

Furthermore this link between the function of a member of a natural kind and teleological explanation makes clear that any Aristotelian account of function must be (in modern terms) evaluatively laden. As Cooper writes: "Aristotle believed that many (not, of course, all) natural events and facts need to be explained by reference to natural goals. He understands by a goal (*hou heneka*), whether natural or not, something good (from some point of view) that something else causes or makes possible, where the other thing exists or happens (at least in part) because of that good" (1982, 197).³ Given this link between function goal and value, it will be clear that, contrary to Wakefield's supposition, an Aristotelian account can give little support to a non-evaluative account of function (Wakefield 2000).

What of Wakefield's other objections? His second objection is that my account provides no basis for the claim that the life of a fully rational animal is the function of a human being. In fact, I give some attention to this point in my earlier paper, and the suggestion I made there is *not* that

attributed by Wakefield (“Megone embraces this argument . . . without further analysis” [2000, 24]), that Aristotle argues by elimination.⁴ Noting Aristotle’s claim that the relevant changes (constitutive of natural function) “come to be always or for the most part,” I point out that “Aristotle is thinking of that regular cycle of changes which contribute to the reproduction of the species and thus to its persistence.” I also note that “Aristotle arrives at this account of the changes that can be explained as the functions of members of kinds by reflection on experience” (Megone 1998a, 193ff.). It may help to elaborate this suggestion a little further here.

The idea is that the identification of the behavior that constitutes the function of a human being presupposes careful and detailed observation of human behavior, so as to be in a position to identify which aspects of the observed behavior contribute to that regular cycle that both ensures and constitutes the persistence of the species. The identification of this cycle is not something that is itself observable, given in the data. It requires reflective judgment on the observed data. In the human case, this judgment is particularly difficult. Since humans are the most complex species, the behavior contributing to the regular cycle is correspondingly complex. Furthermore, as Aristotle observes, humans are a gregarious species so their behavior involves shared projects (a point elaborated below), which adds to the complexity of identifying the relevant cyclical behavior of individuals.

Thus when Aristotle claims that the human function is a fully rational life, he means by this claim to identify a whole cycle of behavior constituting a life. As has been indicated, it is a claim arrived at by reflecting on the range of human behavior we can observe, judging what aspects of the behavior contribute to the regular cycle that is optimal for the persistence of the species and then analyzing the fundamental nature of such a life. In talking of a rational life, he is identifying the function of the whole organism and its function is revealed in a whole life. (This conflicts with Wakefield’s view that the life of reason is the function of a part [2000].)

Furthermore, in characterizing the human function in these terms, Aristotle must have in mind a

pre-theoretical notion of rationality that is in principle of sufficient richness to characterize the whole complex cycle of a good human life, but whose elaboration is to be achieved by reflection on the wide variety of human behavior that can be observed, from which the nature of that relevant cycle is to be extracted. Part of assessing both what rationality involves and whether rationality is the appropriate concept to characterize the function of a human being *qua* human being is reflectively assessing whether those features that cyclically contribute to the persistence and reproduction of the human species are indeed part of a fully rational life. In assessing possible components of such a life, it should be noted that some features may be part of the fully rational life in that they are necessary developments if a human is to become fully rational, while others may be features that are constitutive of rational action and thought.

It follows from this that Wakefield is wrong to suppose there is a conflict between the life of reason and “many features . . . that are developmentally generated as an intrinsic part of the human life cycle and that characteristically contribute to a reproductively successful human life. . . . for example, empathy with others, the sense of justice, complex social emotions like honor and pride and envy, capacity for language . . . ” (2000, 24). These are not competitors to count as the function of a human. On the contrary, these are clearly what Aristotle has in mind in *NE* and *Politics* as part of what a suitably rich conception of the fully rational life will include.

In sum, Aristotle tells us quite enough to make clear how one can set about determining what the human function is. A grasp of this method helps to clarify what is meant by his claim that the human function is a fully rational life (the life of a fully rational animal). In particular it makes clear that this claim can be understood only if it is recognized that he has in mind a very rich conception of rationality exhibited by a language-speaking, emotionally complex, virtuous, social animal (and an animal also exhibiting the traits of theoretical rationality).

Two further points need to be made in elaborating how, on this account, Aristotle identifies

the human function. First note two general assumptions that Aristotle makes regarding the natural world. Aristotle assumes that the natural world is a stable persisting ecosystem (Cooper 1982, 202–5). Second he must suppose that understanding at least some of the behavior of natural kinds as teleologically explicable, involves grasping that a stable ecosystem in which species persist is simply better than a degenerative chaotic system. Presumably this judgment is buttressed by pre-theoretical ordinary beliefs, still widely held, as to how good plants or good animals will develop (Megone 1998a, 193–94). The second point lies at the heart of the way in which the Aristotelian approach to illness conflicts with both that of Szasz and that of Wakefield. The account of how Aristotle identifies the human function is indicative, at least, of one striking way in which the Aristotelian account of the natural world fuses facts and values (at least values of one sort). As was explained in my previous paper, the potentialities that are realized (actualized) when a member of a natural kind fulfills its function also constitute the essential potentialities of members of that kind. Thus that cycle of changes that constitutes the function of the human being as a whole (and the basis for determining the functions of parts) also determines what a human being is. As a result, making the supposedly evaluative judgment as to what constitutes an ill or healthy human being is the very same thing as making the supposedly factual judgment as to what constitutes a human being. The judgments are the same. There is no separation of fact and value.

For this reason, Wakefield's assertions that my view "implies that to call a condition a medical disorder is merely or primarily to make a value judgment and not to make a scientific or factual judgment" (Wakefield 2000, 28) and that on my account, "judgments about what are functions and failures of function are essentially value judgments" (41) are quite wrong. These judgments about functions are both factual judgments and value judgments, insofar as that terminology is applicable to Aristotle. By the same token, Wakefield is wrong when he supposes that the Aristotelian account involves deriving

"the value component from the factual reproductive component in [the] analysis of *function*" (2000, 33). There are not distinct facts and values here. Wakefield, like Szasz, helps himself to the presupposition that facts and values are clearly and necessarily distinct. Though common, that supposition is highly controversial, and it is not applicable to Aristotle's conception of the natural world.

There is a final point to make about the identification of the human natural function. Wakefield himself attributes to Aristotle the view that the fully rational life is the human function but supposes this can be done by "supplementing" the natural kind interpretation of the function argument (2000, 25–28). Some of the remarks already made should have indicated how the natural kind view might identify rationality as the human function. However, it should be noted that the natural kind interpretation can also accept the connection between a rational life and life in a *polis*. Aristotle's remarks on the subordination of ethics to political science (*NE* I.1–2, X.9), and his claims that man is by nature a social animal and that the state is natural (*Politics* I.1–3, 1253a2–5), need to be understood as asserting that the teleologically explicable aspects of human nature can only be fully realized (actualized) in a community. This is also what Aristotle has in mind in characterizing humans as biologically/politically gregarious animals (Megone 1996; 1998c). Wakefield is correct about the connection between reason and community but wrong to think it undermines the natural kind interpretation.

Wakefield's two less serious objections remain to be addressed. First there is the claim that the natural kind interpretation of the argument does not do justice to those aspects of the argument that do not refer to natural kinds. There are three relevant remarks. Two have been dealt with. Aristotle's first point concerning flute players and sculptors (*NE* I.7, 1097b25) introduces a general claim relating the function of a thing to what is good for it. Such a general claim can then be extended to natural kinds, which is what the function argument as a whole aims to achieve. So this first remark can be satisfactorily accounted

for. The two other remarks express reasons why a human being as such might be thought to have a function (*NE* I.7, 1097b28–32). The claim based on the fact that parts of the body are agreed to have functions has been discussed. It has argumentative force *given* the suppressed premise that the functions of wholes are prior to the functions of parts, so that parts only have functions if the whole to which they belong has a function. But what is to be made of the third comment, to the effect that if carpenters and tanners have functions, then so also must human beings? Wakefield is right to suggest that all the remarks not about natural kinds should be accommodated by a plausible account of the argument as a whole, so this remark too should make a contribution to the argument. However, his own suggestion here is not plausible. He takes the point that professions (nonnatural kinds) have socially determined goals (which define them) to support the view that a natural kind such as human being should have the function of a rational life, where the concept of reason is understood socially (that is as a capacity realized in a community). This is a non sequitur.

However, the passages Wakefield cites in discussing this remark are helpful to understanding Aristotle's point here (*NE* I.1, 1094a8–17; I.2, 1094a25ff.). In these passages, Aristotle is arguing that while activities like those of carpenter and tanner are widely believed to have goals (functions), their goals are not goods in themselves. Tables and shoes are desirable for the sake of some further end, means to a further goal. Thus their functions, or goals, are dependent on there being some further goal (at least one) that is a good in itself. His implicit suggestion in this remark is that if humans *qua* humans had a function or goal, that would be a suitable end in itself toward which the goals of socially defined activities, such as carpentry, might lead. In the earlier passages, he intimates that it is the role of political science to investigate this goal (which is consistent with our observation above that the human good or function, the fully rational life, is realized in a *polis*). This argument for the existence of a function for humans *qua* humans is exactly parallel to the argument that the goals of

parts of the body are subordinate to the goal of the human being as a whole. Thus here all the goals of subordinate artificially defined activities are subordinate to the function of humans as such. This reference to socially defined activities does then provide a further reason for supposing that humans as such have a function. But the form of the reasoning here does nothing to undermine the natural kind interpretation of the function argument.

Finally, Wakefield's fourth objection is that the natural kind interpretation renders the function argument unhelpfully circular in an investigation of the human good. For, he suggests, analyzing the function of a human being in terms of some goal or purpose that teleologically explains certain behavior makes the argument circular since we have to appeal to the notion of the good human in order to specify what behavior is functionally explicable (Wakefield 2000, 28–29). What is correct here is that Aristotle is *not* arguing in favor of identifying the human function first, so that the human good may be determined on that basis (as Wakefield seems to think he is). The human function is that aspect of human behavior that is teleologically explicable, and the concept of the good human is relevant to identifying the teleologically explicable behavior (Megone 1998a, 193). However, Aristotle does not present us with a trivial circle, in which the human function just is what the good human does, and the good human just is one who performs the human function. Rather than this, Aristotle's suggestion that the human function might cast light on the human good introduces to an investigation of what is good for a human a much larger picture of the natural world as teleologically explicable. With this much larger picture comes a corresponding understanding of the role of a good member of the species in that world, which can be used to identify concretely which aspects of human behavior a good human would perform. Each of these conceptions, the teleological conception of the natural world and the conception of the good human, is developed in tandem; and each must also be sensitive to a range of other theoretical beliefs (in the case of the natural world, beliefs about natural kinds; in the case of

the human good, other beliefs about what is good for humans, such as beliefs about human psychology and physiology, and about the virtues). The argument may suggest that certain concepts need to be understood together, but not in any manner that is viciously circular.

In this way, the natural kind interpretation of Aristotle's function argument can be defended against Wakefield's criticisms. The basic idea we are left with then is that human beings, like members of other natural kinds, have a function in the sense that some of their behavior, but not all, is teleologically explicable, that is, can be understood as achieving some natural goal or purpose, something that is, from some point of view, good. Such an account of the human function is necessarily evaluative, but it is also necessarily factual since it is tied to what a human being essentially is. Aristotle claims that the human function is, within this account, the life of the fully rational animal. Illness is any incapacitating failure to realize (actualize) this human function.

A NON-EVALUATIVE ACCOUNT OF FUNCTION?

Although the Aristotelian evaluative account of human function, and the corresponding analysis of illness or disease, remains intact, it is worth considering the merits of Wakefield's suggestion that the concept of human function can be analyzed non-evaluatively. What has already been shown, though, is that Aristotle's texts will not provide support for such an analysis.

Wakefield characterizes his claim as the view that *function* and *dysfunction* are factual concepts (Wakefield 2000). Three stages in the defense of this position will be considered. First of all, Wakefield suggests that there are examples of failures of function that are not harmful, indicating that a purely factual understanding of the concept is desirable. Second, he argues that there is an alternative account of *function* within which functional behavior can be seen as teleologically explicable behavior, but teleological explanation itself can be reduced to a special kind of cause/effect explanation and thus need not require any normative framework. Third, he argues that such

a factual understanding of *function* is supported by an analysis of its role in an evolutionary account of natural kinds.

It is worth noting first the way that Wakefield sets up this part of the debate. He claims that I "assume that the value account of function is correct, based on [my] view that past attempts to analyze function entirely in non-evaluative scientific terms have failed" (Wakefield 2000, 28). Wakefield's claim about my motivation here is mistaken. The claim is partly true in that I do take it to be relevant that previous attempts at non-evaluative accounts of function have failed, a point made by Fulford (Fulford 1991, 84, 98). However, I also emphasized that the ordinary concepts of illness and disease are clearly evaluative notions, so that it will be an advantage for a functional account of these concepts if the concept of function is itself evaluative (Megone 1998a, 191). But a second point about Wakefield's remark depends on noting again the assumptions he is making here. As noted above, he assumes that there is an essential contrast between facts and values and thus that any scientific account of illness must be factual and therefore *not* evaluative. As has also been indicated, though, the Aristotelian account does not share this fact-value distinction or, therefore, this view of a scientific analysis of the natural world. Facts and values are fused in the natural world, so that an evaluative account of function may also be a factual account, and such an account may thus be thoroughly scientific. The evaluative account of function, and corresponding account of illness, does not provide an evaluative account rather than a scientific account. It reveals the way in which a scientific account of illness involves a judgment that is at one and the same time both factual and evaluative. Thus Wakefield's view, that he needs a non-evaluative account of function to preserve the scientific status of the concept, is mistaken.

Wakefield supposes, then, that I, like many others (e.g., Sadler and Agich 1995), simply assume that *function* and *dysfunction* are primarily evaluative concepts. What has been said so far in reply is that there are good pre-theoretical reasons to seek such an account, and that there

are theoretical reasons why any such account will not give rise to unscientific concepts. This already rebuts the thought that a mere assumption was being made. In any case, though, the underlying thought in fact arises from reflection on the function of artifacts, where the term has its primary application. An artifact like a knife can do many things in the sense that it can be used to bang a table, hammer a nail, smash a glass, complement a fork, and so on. These are all powers or potentialities of a knife. What determines what potential must be realized if the knife is to perform its function, as opposed to one of the many other things it can do? The answer is that its function is given by its designer, and that is the goal, or good, that the designer aims at in designing the knife. Reference to the good, as perceived by the designer, is essential in picking out the function of an artifact. Correspondingly when we explain why a certain implement is being used on the bread, by citing the purpose of cutting, we are providing a teleological explanation of the activity in which the function of that implement figures as the good that the designer/user aims at in this use. If the case of artifacts is taken as a starting point for an understanding of *function*, it is clear that an evaluative account of the concept falls naturally out of the analysis.

As a result, when the language of function is transferred to the case of members of natural kinds and their parts, it is reasonable to expect the same analysis. Eyes and feet and hearts can each figure (or rather their activities can so figure) in the production of a wide range of effects. If so, why is it the function of eyes to see or of hearts to pump the blood? The parallel with artifacts (a parallel that Aristotle also exploits, as noted, [*Physics*, II.8–9]) suggests that these particular effects, produced by the operation of eyes and hearts in certain circumstances, only constitute their function because there is at least some perspective from which these activities can be seen as good, thus a goal of the organ in question. Thus I did not hold that the function of a heart must contribute to reproduction (or that *function* means “characteristically contributes to reproduction”), as Wakefield reports (2000, 29). The fundamental idea is that the heart’s function

must be a goal, good from some perspective, and pumping the blood can be seen as achieving a goal if that activity contributes to the persistence of the species, and that in turn is, from some perspective, good. I suggested that the sense in which all this is good is given if we recognize that a stable, ordered, persisting ecosystem is better than a chaotic, degenerating ecological environment (Megone 1998a, 194). All this is consistent with the claim made earlier that fundamentally the function of a thing is *that aspect of the thing’s behavior (whether the thing be a whole or a part) that is open to teleological explanation*. What is being brought out here is the way in which that analysis, starting in a natural way from consideration of artifacts, gives rise to the view that *function* is an evaluative concept. This confirms that the treatment of *function* as an evaluative notion was a matter of analysis, not simply an assumption.

Even if not merely assumed, the evaluative account of *function* may be mistaken. Wakefield argues that it is. He claims first that the perspective referred to above, in which a stable, ordered, persisting ecosystem is seen as good, cannot in fact explain why certain activities of natural kinds and their parts constitute the functions of those things. To support this, he asserts that: the parts of viruses would still “require functional explanation” even from those who think it better that such viruses cease to exist; “the parts of a human being would still have functional explanations” even if it were better if the human race had never existed; and the parts of an organism that is designed to destroy ecosystems would be “just as much candidates for functional explanation as those of the most ecologically benign organism” (Wakefield 2000, 30). However, he gives no explanation for this assertion. The evaluative account of *function* only holds that if the activity of a member of a natural kind, or of its part, is to be open to functional explanation, there must be *some perspective* from which that activity is seen as good. Only from that perspective is the behavior functionally explicable. Of course functional explanation of the behavior of parts will be possible in the cases above if it is said, for example, that the survival of the virus, or of humans, or of

the corrupt organism, is good in the sense that it is good, in each case, for that kind itself, or good from the perspective of genetic diversity. (Wakefield queries whether it makes sense to talk in the first way about what is good for a virus) (2000, 32). Presumably we can make some sense of a predator being bad for a species because it threatens the survival of the species. But in any case, here no commitment is made to this being a coherent perspective.) There will often be a range of perspectives on the value of some activity. The point is that functional explanation will be possible to the extent that there is some perspective from which it has positive value.

Of course if in the end, it is held that any or all of these perspectives are *mere* perspectives, that there is no genuine overall good served by the relevant activities, then such explanation will be mere “as if” teleology. There will not, on such a view, be genuine functional explanations. Talk of the function of natural kinds or their parts will be merely a way we have of thinking about the world (Akrill 1982; Wieland 1975).

Wakefield’s mere assertion that it will still be possible to give functional explanation of the activity of parts in the cases he mentions leaves it unclear whether he thinks that such explanations are merely ‘as if’ teleology, or whether he thinks there are other goals (other than a stable ecosystem) that are from some perspective genuinely good, and which these activities serve. If he holds neither of these views his claim that functional explanation is possible, even if the survival of a species is a bad thing, is a mere assertion.

Compare a clock. If we learn that an artifact we had supposed to have the function of telling the time was not made by its designer to achieve that goal, then it is no longer the case that the hands have the function of showing the time, at least unless there is some other perspective of the designer’s from which such activity is good. If we learn that there is no perspective from which the behavior of the artifact is good from the point of view of the designer (or user), this undermines the claim that the parts have any function.⁵ So the claim that the activity of an organism constitutes the function of an organism requires only that the activity be good from *some* perspective.

Sensitivity to this nuance explains why Wakefield’s supposed counter-examples appear to have any plausibility (2000, 19, 30–32). The examples trade on the reader viewing them from such perspective. But in fact, what Wakefield needs is the claim that the parts’ behavior will clearly be functionally explicable even if no such perspective is possible. No argument is given for such a claim in any of the cases given.

This can be illustrated with reference to just one of the examples given. Wakefield picks out the thesis that social welfare policy has the function of controlling the poor rather than of helping them (2000, 32). He argues that this shows that functional explanation need not be evaluative since this is, he claims, a perfectly valid functional explanation; yet, none of the authors who offer such a functional explanation of social welfare policy think that it is a good thing that the poor are controlled in this way. But the evaluative account of functional/teleological explanation holds only that if some activity is functionally explicable, that activity must be so explicable by reference to a goal that is, from some point of view, good. In this case, social welfare policy is functionally explicable in virtue of a goal that is (according to a certain theory) good from the perspective of the upper classes. There is no requirement that one who cites such an explanation endorse that perspective. However, Wakefield trades on the fact that the reader accepts the availability of that perspective in order to make it plausible that functional explanation is possible here. Were no such perspective possible, social welfare policy would not be functionally explicable.

Quite generally, one who cites a functional explanation alludes to the existence of a relevant evaluative perspective but need not endorse it. Similarly, to accept that an activity is functionally explicable by reference to a goal that is good from some perspective is not to commit the theorist to the view that the goal is all things considered good, or exceptionlessly good.⁶ As has been said, sensitivity to this nuanced account allows all Wakefield’s purported counter-examples to an evaluative account of function to be handled in ways similar to that above.

However, Wakefield has a theoretical objection that goes beyond this appeal to purported counter-examples. He claims that “the addition of such a value component does not warrant teleological explanation. The eyes have the effect of seeing, and seeing is characteristic of a human life that leads to reproduction, and we value the good life of which seeing and reproduction are a part, but these facts together do not contain a distinctive teleological element. . . . Nothing in this combination of facts and values warrants teleological explanation. So the basic problem is this: When a feature causes a valued effect, that is still just standard efficient causation plus a value judgment about the effect. It is not a distinctive form of teleological explanation in which the effect can be used to explain the feature.” Wakefield then proposes an alternative:

Rather than holding that a mechanism’s effect is a function and can be used to teleologically explain the mechanism when we value the effect, one can hold that the effect is a function when it must be cited in a causal explanation of the mechanism, as in an explanation by natural selection in which the effect of a mechanism explains why [it] exists and is maintained in the species and why it is structured the way it is. Note that such explanations are merely an unusual form of efficient causal explanation. Thus, this analysis dissolves the central mystery about teleological explanations; namely, why do they *explain* anything? Citing a valued effect does not offer any explanation whatever of a mechanism, but citing an effect that played a role in the causal sequence leading to the mechanism is clearly explanatory in a sense with which we are familiar. (2000, 31)

Such a long quote is necessary both because this passage lies at the heart of Wakefield’s defense of a non-evaluative account of *function* and because the passage is rather curious. Does Wakefield have any argument here for the view that explaining the behavior of some whole or part by reference to the goal of that activity is not a genuine, *teleological* explanation? There are at least four problems with what appears to be the line of argument.

Perhaps the most fundamental of these problems is to be found toward the end of the passage. Wakefield seems implicitly to hold a theory of explanation according to which only cause and effect explanations are genuinely explanatory (at

least of mechanical behavior). The holding of such a theory would also account for another rather curious remark: “If the value theorist concedes that an explanatory component is essential to the meaning of *function* . . . ” (Wakefield 2000, 33). This is an odd remark since the value theorist sees the function of a structure as its teleologically explicable behavior, so such a remark could only make any sense to someone who supposes that explanation citing valued objectives is not genuine explanation and that only cause and effect explanation is genuinely explanatory. But Wakefield gives absolutely no argument for this idiosyncratic view of explanation. In his analysis of the nature of explanation, Aristotle provides us with four different explanatory modes, formal, material, agent–patient (which is not in fact quite the same as the Humean cause–effect mode, and may have advantages over it), and teleological (*Physics* II.3, 194b16–195a4). These are four quite distinct schema, each of which provides an irreducibly different mode of understanding of the world. Formal, material, and teleological are not reducible to agent–patient explanation, but nonetheless they are required for a full understanding of the world. The formal explanation for a figure’s being a square, that it has four equal sides and four right angles, is genuinely explanatory but not reducible to cause–effect explanation. The material explanation for a statue’s being heavy, that it is made of gold, is genuinely explanatory, but not reducible to cause–effect explanation. So, finally, the teleological explanation for a knife’s being sharp, that it is for cutting, or for a fish’s having fins, that they are for swimming, is genuinely explanatory but not reducible to cause and effect explanation.

Wakefield seems to provide no argument for rejecting the explanatory power of these distinct modes of explanation. He simply asserts it. Yet on the face of it, only such a belief can make sense of his denial that to explain the existence or behavior of eyes by reference to seeing, where vision is conceived as a goal, from some perspective good, is to provide a genuine teleological explanation. More than assertion is required to support such a denial. For it is very odd to think that this

mode of explanation can only be genuinely explanatory if it can somehow be reduced to cause-effect explanation.

In fact, teleological explanation provides a further level of understanding of the world over and above cause-effect explanation. In some cases, in addition to being able to explain *how* it comes about that an agent is engaged in the activity of seeing, by citing the causal chain that is bringing about that activity, it will *also* be possible to explain that activity as being (or producing) a goal, good from some perspective. Once such a teleological explanation is grasped, it will be possible to understand not just how an activity has come about but also that, since the activity has come about, things are, from some perspective, going well (or badly, if the activity constitutes a failure of a goal to come about). This is a distinct mode of understanding of the world. Obviously it is particularly important in the field of medicine since it is crucial to know not just how a condition has come about but whether that condition's coming about constitutes things going well (health) or badly (illness) for the organism. Only in the light of the understanding given by this teleological mode of explanation is it possible to determine whether medical intervention is or is not necessary. In other spheres, such as mathematics, the teleological mode of explanation may not be applicable (just as the cause-effect mode may not be). So there is every reason to think teleological explanation is genuinely explanatory, even if not reducible to efficient causal explanation. If so, Wakefield's apparently suppressed general view of explanation can be rejected.

Two further problems in the passage outlined above now follow. First, Wakefield's attempt to cite the evolutionary account of natural selection in support of his view seems questionable. Second, his positive alternative account of teleological explanation is open to doubt. On Wakefield's view within evolutionary theory, the claim that the heart has a natural function (pumping the blood) should be seen as specifying the special causal role of the heart's pumping of the blood. This suggestion can be addressed directly. However, it is at least as plausible to view evolution-

ary theory differently. The theory is not wholeheartedly teleological in that it supposes that the actual occurrence of entities belonging to species that will successfully survive is to be explained through the chance interaction of physical entities. However, once a member of a species exists, if within evolutionary theory, parts of that organism are said to have a natural function, a natural construal of this claim is that the behavior of those parts can be explained in a way that is irreducibly teleological. That is, when evolutionary theory holds that hearts exist for the purpose of the pumping of the blood, it implicitly treats survival of the species as a goal, from some perspective good, and thus treats pumping of the blood as a derivative goal, derivatively good from that perspective. (This point has at least been raised by Agich and Sadler in discussing Wakefield's position (Agich and Sadler 1995, 224–27).

In any case, one can question whether Wakefield's proposed alternative account of teleological explanation applies satisfactorily either to this case or to the functional/teleological explanation of artifacts. According to Wakefield, a teleological explanation is provided when there is a causal explanation of an effect and the effect is itself the cause of the existence of the mechanism that causes it. In this case, "the effect is a function" (Wakefield 2000, 31). The first problem with this account of teleological explanation is that it loses any sense in which this is explanation by reference to a *telos*, or goal or purpose, so it is unclear why it is thought to be teleological.

The second difficulty is that it is very doubtful that this structure can be made to apply to either biological or artifact cases. According to Wakefield, on this model "the heart exists for the purpose of pumping the blood in the sense that the fact that past hearts had this effect causally explains how hearts came to exist and be maintained in the species and the genesis of the heart's detailed structure" (2000, 31). The most basic difficulty here is that past hearts' pumping the blood figure in all sorts of causal stories, stories in which agents die young, or agents do not reproduce, or agents reproduce but defectively, and so on. It is not the case that hearts pumping the blood have

simply caused hearts to exist. So hearts pumping the blood does not causally explain how hearts came to exist. This cannot therefore be the sense in which the pumping of the blood is a functionally explicable activity of the heart.

The best that might be said is that there is one causal chain, out of the myriad causal chains in which the heart's pumping figures, in which, along with the operation of the rest of the body, does cause another heart to exist. But it is plausible that the only way to distinguish this causal chain from all the myriad others is to say that this is the chain in which the heart's pumping achieves its goal, survival of the species, which makes the pumping here thus derivatively good. The relevant causal chain is the one in which the pumping is teleologically explicable in the goal-based sense. In other words, this approach is led back to the teleological mode of explanation in the standard Aristotelian sense. The problem of applying Wakefield's suggestion to the artifact case is worse still. According to the suggestion, cutting is the teleologically explicable activity of a knife in the sense that the effect, cutting, "was anticipated in the [mind] of the [designer] and played a causal role in how [he] designed the artifact" (Wakefield 2000, 31). The difficulty is that here it is not the effect, cutting, that plays *any* causal role in the design of the artifact. What may play a causal role (though even that is disputed) is the designer's *desire* for a cutting implement. But in any case, that desire for a cutting implement once again figures in all sorts of causal stories, stories in which the agent does intend to produce a knife, or does not, and does, or does not, achieve his intention. The desire for a cutting implement does not simply cause a cutting implement to exist. This cannot therefore be the sense in which cutting is a functionally explicable use of a knife.

There is (at least) one causal chain, out of many in which it figures, where this result, cutting, does occur. However, again it is plausible that in order to pick out this chain (or these chains) as the one that is relevant to its functional role, it will be necessary to specify this as the chain that brings about the goal of the desire (which the designer thought of as good). Once

again the standard Aristotelian account of teleological explanation must be appealed to in order to rescue the supposed alternative account.

In sum, Wakefield admits that functionally explicable activity of a whole or a part is teleologically explicable activity. However, he cannot produce plausible examples of teleologically explicable activity that is not, from some perspective good. Nor can he produce a satisfactorily non-evaluative account of teleological explanation. So he cannot find any support here for a non-evaluative account of the concept of *function*.

The discussion has also identified a potential underlying problem. Wakefield's efforts to produce a non-evaluative account of function seem driven by a mistaken set of interrelated suppositions: that there is a sharp fact/value distinction; that science is concerned only with facts; that a genuinely scientific (and genuinely explanatory) explanation of the world must therefore be purely factual; and that cause and effect explanation constitutes such a mode of explanation. In response to this set of suppositions, the Aristotelian approach presents a picture of the natural world in which facts and values are fused, and thus a correspondingly richer account of modes of explanation, within which to locate an evaluative account of the concept of *function* and corresponding function based account of illness.

Wakefield's account of illness concedes that both functional norms and values are relevant to an adequate account of the concept. In a similar way, Szasz, too, has conceded a role for both the notion of function and that of value. To this extent, both provide support for an account of the concept in terms of an evaluatively understood concept of function. The failure of Szasz to provide an alternative source for functional norms to that given in the Aristotelian account and the failure of Wakefield to provide a non-evaluative alternative to the evaluative account of function leave the Aristotelian account as the best account, compatible with the points they have both conceded, that is still in the field.

* * *

PRACTICAL IMPLICATIONS OF THE ARISTOTELIAN FUNCTIONAL ACCOUNT OF MENTAL ILLNESS

The Aristotelian account of the function of human beings may offer the beginnings of a better understanding of mental illness. It is only the beginning since, on this account, illness is an incapacitating failure of a part of a human to act as it must if the agent as a whole is to live a fully rational life (strictly, the life of a fully rational animal). This account is incomplete since it leaves for further elaboration the question of what the concept of rationality here involves, although some indication of the nature of the conception has been given and some concrete features have been indicated.

However, Szasz reiterates in his response that such a philosophical analysis is of no relevance since it ignores the fact “that psychiatry is, in effect, a branch of the law” (2000, 9). His objection appears to be that such an analysis is of no practical use since it can have no positive effect on the practices in which psychiatry is, in his view, currently involved. This point has already been addressed in a previous response, so I shall be brief (Megone 1998b, 223).

Szasz elaborates this claim about psychiatry by citing a number of recent cases in which people have been, according to his report, abused by psychiatrists. He claims that this illustrates “the truth . . . that, in the United States alone, each year hundreds of thousands of persons are subjected to psychiatric coercions” (2000, 12). In his view, these cases reflect the way in which “[psychiatry’s] institutions and interventions legitimize [coercive] relations between rulers and ruled . . . ” (2000, 15). According to Szasz, I imply “that [such] compulsion is rarely used in psychiatry” (2000, 12), and my whole approach fails to address the way in which psychiatry deploys the mythical notion of mental illness in order to legitimize an indefensible ideological political order.

Szasz seems still not to appreciate that the Aristotelian analysis of mental illness is intimately concerned with the legitimacy of psychiatric practices. Such concern is not only manifested in

endorsements of Szasz’s own position. Whatever one’s position on the matter, an understanding of the nature of mental illness is an essential prerequisite to both understanding and critically assessing the social and legal practices governing the treatment of the mentally ill. Viewing the issue from a very general perspective, the difference in the way in which mental and physical illnesses incapacitate an agent in rational functioning might be expected to distinguish the way in which the mentally and the physically ill are treated. Very roughly, physical illnesses incapacitate rational functioning by indirect attacks on the agent’s rational powers, either preventing rational choices being acted upon (too sick to run, for example), or preventing receipt of the information on which such choices depend. Mental illnesses, on the other hand, incapacitate rational functioning by direct attack on the agent’s rational powers, preventing the agent from forming rational beliefs on the basis of information or from making rational choices in the light of desires and beliefs. It will therefore not be surprising if the whole mode of treating the mentally ill, incapacitated from making choices, for example, differs from that of agents who, through physical illness, are incapacitated from acting on choices. This general picture thus tends to endorse the practical implications of the Aristotelian perspective as against Szasz’s.

To avoid too much repetition, I will simply summarize the differences between my approach and Szasz’s.

1. In my paper, I was not analyzing any actual cases to determine the legitimacy of coercion on those occasions, and Szasz is quite wrong to construe my claim that “[c]ompulsory treatment can be justified at least for some patients” as in any way implying that compulsion is rarely used in psychiatry.

2. What that claim does assert is that compulsory psychiatric treatment can *sometimes* be justified. This is because, in my analysis, mental illness is not a myth. On the contrary, mental illness occurs when an agent is incapacitated from rational belief formation or rational choice. Coercion *may* be justified in such a case since a patient who is ill in such a way lacks a basic

requirement for the realization of autonomy. Hence “compulsory treatment of mental illness, rather than infringing autonomy, may in fact facilitate its recovery” (Megone 1998a, 199).

3. This analysis does not hold that all psychiatric coercion is unjustifiable, yet it is also able to explain the possible misuse of psychiatry. The Aristotelian account of mental illness incorporates the notion of rationality, and, as has been seen, this concept is, on that account, immensely complex and difficult to spell out. It is this fact and the association or practical rationality with goals that “has made it possible (but not defensible) for some highly disputable norms to be imposed” (Megone 1998a, 199).

4. Such a position, while not endorsing Szasz’s extreme view that mental illness is a myth, does endorse the view that “caution is required in the attribution of mental illness” (Megone 1998a, 199). This point will be elaborated somewhat below.

5. This account of mental illness via a functional analysis does not account for the phenomenon of illness in purely physico-chemical terms. In the absence of such a reduction, “the possibility remains open that some illnesses, at least, may not be best treated physico-chemically” (Megone 1998a, 199). Thus wariness about drugs treatment is vindicated, but the physical realization of conditions constituting illnesses also leaves it open that drugs could sometimes be of benefit.

One further point is worth emphasizing again. It is a widely held pre-theoretical belief that mental illness does occur. The possibility of explaining practical phenomena associated with it without being driven to denying its existence is a theoretical advantage of the Aristotelian account over Szasz’s position.

DEVELOPING THE ARISTOTELIAN ACCOUNT

Considerable work remains to be done to develop the Aristotelian account. The rich account of rationality requires articulation in order to identify with much more precision possible criteria for identification of specific mental illnesses. In fact, these projects are mutually compatible in that proposed accounts of mental illnesses can be drawn upon in order to advance proposals for a

deeper understanding of rationality and, at the same time, proposals regarding rationality can be used to clarify conceptions of mental illness. This is a further clear indication of the way in which work in philosophy and work in psychiatry will be of great mutual benefit (and of the way in which concepts can be clarified together without circularity).

It will also be necessary to distinguish mental illnesses from other conditions that lead to functional failure. As was explained previously, laziness can lead to functional failure, just as mental illness can (Megone 1998a, 196), and in Aristotle’s account, moral vices will constitute irrationality, as will schizophrenia. The theoretical distinction between laziness, or any moral vice, and mental illness is that the former is within the agent’s power (*hekon*, in Aristotle’s terminology) while the latter is an incapacitating failure (*akon*, in Aristotle’s terminology) (*NE*, III.1–5; Hobbs 1998, 211–12). However, initially at least, the epistemological criteria for distinguishing these conditions or, in general, for distinguishing mental illnesses from other conditions in which irrationality is exhibited will be behavioral. It may be that in due course there will in some cases be support for the identification of some conditions as indeed illnesses (incapacitating conditions) from neurological evidence, but the present level of neurological knowledge means this will tend to be supportive, at best, at present.

The difficulties in making these distinctions again seem perfectly compatible with widely held pre-theoretical beliefs as to, for example, the difficulties in determining whether some agents of criminal activity should be deemed mentally ill or judged responsible for their acts. Thus the Aristotelian functional theory of mental illness once again meets the theoretical criteria for a good account.

CONCLUSION

Though the aim of this paper has been defensive, it is to be hoped that the range of issues raised by the forthright challenges from Wakefield, Szasz, and previous commentators has made it possible for its contribution to be positive as well. In particular, this defense of the Aristote-

lian response has served to highlight a number of the key presuppositions that inform the accounts of Wakefield and Szasz, on the one side, and their Aristotelian challenger on the other. Despite their differences, both Szasz and Wakefield appear to share a conception of science as an empirical discipline concerned with the material world. In this view, science is concerned with facts, and facts are given in observation, and what is observed is the material realm. There is a sharp contrast between facts and values. Values are not observable, nor material, and thus not the concern of science. To the extent that illness has anything to do with values, therefore, that evaluative aspect of the concept has a non-scientific source (though, as has been indicated, Szasz's position on this last matter is less clear). Two further presuppositions are compatible with this materialist empiricism. On the one hand, there is Szasz's Cartesian dualism about the mind, treating the mind as an immaterial substance and thus perhaps beyond the concern of science. On the other hand, there is Wakefield's theory of explanation that views only efficient causal explanation found, he may suppose, in the material realm, as the genuinely explanatory mode of explanation to be given in science. Their accounts of the concept of mental illness are built on these presuppositions.

Contrary to this metaphysical scheme stands the Aristotelian framework in which the scientific investigation of the natural world is an investigation of a world in which facts and values are fused. Correspondingly, the Aristotelian account rejects the fact/value distinction, rejects a Cartesian substance dualism about the mind and the body, and offers a much richer scheme of genuinely explanatory modes of (scientific) explanation that accords with the nature of the natural substances that constitute the world that science investigates. It is not surprising that an Aristotelian account of illness, which is a facet of a natural kind (in this case, human illness), should therefore reflect this fusion of facts and values and the important role of irreducible teleological explanation within this metaphysical picture.

The discussion has also touched on underlying issues concerning the nature of the meaning (or

definition) of concepts and on method in philosophical inquiry. These are relevant to the development of the metaphysical pictures needed to understand the concept of mental illness. However, I hope it has become clear that it is in the distinction between the Aristotelian metaphysical picture and that of its opponents that the contrast between the different positions is sharpest and that there is much to be said for the Aristotelian position.

NOTES

1. Szasz's position has been briefly addressed previously (Megone 1998b, 223), but the analysis that follows may make the rebuttal clear.

2. As a matter of fact the *OED* gives a definition of illness as (1) disease, (2) the state of being ill (which conforms with Szasz's treatment of disease and illness as interchangeable), and defines ill as out of health, sick (. . . *mentally ill people*), so Szasz does not receive dictionary support here. But my aim is not to challenge him on these grounds.

3. Cooper cites *Physics* II.2, 194a 32–33; II.3, 195a23–25; *Politics* I.2, 1252b34–35, and *EE* I.8, 1218b9–11 as among “the many passages where Aristotle routinely explicates ‘that for the sake of which’ by linking it with the good, the fine, the better, etc. (cf. *MP* 13, 983a31–2, *PA* 11 639b19–20). That the concept of a goal is the concept of something good is a view Aristotle inherited from Plato's *Phaedo* (cf., for example, 97C6–D3, E1–4, 98A6–B3, 99A7–C7)” (Cooper 1982, 197).

4. Wakefield seems simply to have overlooked or ignored the fact that the question of how the function of a kind is to be determined is specifically raised and addressed in my first paper (Megone 1998a, 193ff).

5. The example would now become problematic if one wondered whether it entailed an agent making something to no purpose whatever (Anscombe, 1957). Presumably the case would have to involve an agent simply manipulating parts, the mechanical equivalent of doodling, and thus the parts have no function.

6. For example, a certain activity may be good at many times in the history of a species, but not at all times; so it is not an exceptionless good.

REFERENCES

- Ackrill, J. 1982. *Aristotle the philosopher*. Oxford: Oxford UP.
- Agich, G., and J. Sadler. 1995. Diseases, functions, values, and psychiatric classification. *Philosophy, Psychiatry, & Psychology* 2, no. 3:219–31.
- Anscombe, G. E. M. 1957. *Intention*. Oxford: Blackwell.

- Aristotle. 1992. EE. *Eudemian ethics, books I, II, and VIII*. Trans. M. Woods. Clarendon Aristotle Series. Oxford: Oxford UP.
- . 1980. NE. *Nicomachean ethics*. Trans. W. D. Ross. Rev. J. L. Ackrill and J. Urmsen. Oxford: Oxford UP.
- . 1992. Phys. *Physics, books I and II*. Trans. W. Charlton. Clarendon Aristotle Series. Oxford: Oxford UP.
- . 1995. Politics. *Politics*. Trans. E. Barker. Rev. R. F. Stalley. Oxford: Oxford UP.
- Boorse, C. 1975. On the distinction between disease and illness. *Philosophy and Public Affairs* 5:49–68.
- Cooper, J. 1982. “Aristotle on natural teleology.” In *Language and logos*, ed. M. Schofield and M. Nussbaum, 197–222. Cambridge: Cambridge UP.
- Fulford, K. W. M. 1989. *Moral theory and medical practice*. Cambridge: Cambridge UP.
- . 1991. The concept of disease. In *Psychiatric ethics*, ed. S. Bloch and P. Chadoff, 2d ed., 77–99. Oxford: Oxford UP.
- Hobbs, A. 1998. Commentary on “Aristotle’s function argument and the concept of mental illness.” *Philosophy, Psychiatry, & Psychology* 5, no. 3:209–13.
- Kendell, R. E. 1975. The concept of disease and its implications for psychiatry. *The British Journal of Psychiatry* 127:305–15.
- Macklin, R. 1973. The medical model in psychoanalysis and psychotherapy. *Comprehensive Psychiatry* 14:49–69.
- Megone, C. 1996. Critical discussion of *A companion to Aristotle’s politics*. Ed. D. Keyt and F. Miller. *Polis* 14:165–80.
- . 1998a. Aristotle’s function argument and the concept of mental illness. *Philosophy, Psychiatry, & Psychology* 5, no. 3:187–201.
- . 1998b. Response to the commentaries. *Philosophy, Psychiatry, & Psychology* 5, no. 3:221–24.
- . 1998c. Aristotelian ethics. In *The encyclopaedia of applied ethics*, vol. 1, ed. R. Chadwick, 209–32. San Diego: Academic Press.
- Szasz, T. 1960. The myth of mental illness. *American Psychologist* 15:113–18.
- . 1998. Commentary on “Aristotle’s function argument and the concept of mental illness.” *Philosophy, Psychiatry, & Psychology* 5, no. 3:203–07.
- . 2000. Second commentary on “Aristotle’s function argument.” *Philosophy, Psychiatry, & Psychology* 7, no. 1:3–16.
- Wakefield, J. C. 1995. Dysfunction as a value-free concept: A reply to Sadler and Agich. *Philosophy, Psychiatry, & Psychology* 2, no. 3:233–46.
- . 2000. Aristotle as sociobiologist: The “function of a human being argument, black box essentialism, and the concept of mental disorder.” *Philosophy, Psychiatry, & Psychology* 7, no. 1:17–44.
- Wieland, W. 1975. The problem of teleology. In *Articles on Aristotle*, vol. 1, ed. J. Barnes, M. Schofield, and R. Sorabji, 141–60. London: Duckworth.