

Functionalism and Moral Personhood: One View Considered

DAVID C. WILSON

University of California, Los Angeles

In *Brainstorms*, Daniel Dennett sets forth a series of necessary (and almost sufficient) conditions for admitting a being into the moral community — conditions which are, for the most part, pragmatic in nature. We are to consider an entity as the proper object of moral commitment (and thus as a person) just in case it is *useful* to us to view that being in certain ways. Although there are many advantages to such a strategy, Dennett develops it in such a way that the list of those beings he would ostracize and those he would welcome differs significantly from our intuitive notions. In this paper I will summarize the pertinent features of his view and some of the unpalatable consequences.

1. Dennett's View

It is only when we adopt what Dennett terms alternately the “personal stance” or the “moral stance” toward an entity that we can consider it to be free, responsible, and the proper object of moral commitment (a notion which is never given any greater precision). He explicitly offers four necessary conditions for the “moral personhood” of an entity, requiring that:

- (i) it is an intentional system,
- (ii) it is capable of reciprocity,
- (iii) it is capable of communication, and
- (iv) it is conscious (or self-conscious).

Although he is hesitant to consider these conditions jointly sufficient, they seem to function in Dennett as sufficient conditions, at least for practical purposes. Some elaboration of these conditions will help fill in the picture.

Intentional systems. Condition (i) is the most fully developed by Dennett. He declares repeatedly that we are to ascribe intentions (i.e., beliefs, desires, etc.) to an object just in case (a) our goal is to explain and predict

its behavior, and (b) this goal can be achieved successfully (and perhaps *most* successfully — see pp. 238, 271)¹ by assuming the object's rationality. In his words, intentionality "is a matter settled pragmatically, without reference to whether the object *really* has beliefs, intentions, and so forth" (p. 238). By assuming the object's rationality, one has adopted what Dennett refers to as the "intentional stance," and the object is ipso facto an "intentional system."

In those cases when rationality is not a useful assumption, we may turn instead to either Dennett's "design stance" or his "physical stance." The former we are to adopt when it is most useful to understand the being according to its *functional structure* (e.g., a computer's program), while the latter we should employ when a description of its *physical structure* (e.g., a computer's wires and switches) best suits our needs. It should be remarked, however, that these two stances have an ontological commitment, while the intentional stance is adopted, as noted above, "without reference to whether the object *really has beliefs*. . . ." In fact, he frequently expresses doubts as to whether such entities exist at all; for example, "eliminative materialism" is a designation that he believes best typifies his position, since, except for functional notions, "we legislate the putative items right out of existence" (p. xx; see also p. xix and chapters 8 and 10).

Dennett's other explicit conditions. Now, (i) is necessary for all three remaining conditions; furthermore, (ii) (reciprocity) is a necessary condition for (iii) (communication), while (iii) is both a necessary *and sufficient* (see p. 271, n.) condition for (iv) (consciousness). Each of these three is like (i) in that it is to be attributed to a being just in a case it is useful to do so, without regard to whether some corresponding property is actually present.

This clearly follows for Dennett, given that he understands each condition merely as reflecting *additional levels of intentionality*. Reciprocity is a matter of the being having second order intentions (i.e., ascribing to it *beliefs* about *beliefs*); communication is primarily a matter of third order intentions (ascribing to the system the *intention* that I *recognize* that it *intends* that I respond in a certain way); and consciousness involves second order intentions again (attributing to the system the *desire* to have a *desire*). Clearly, Dennett would not be consistent if he did not require the same pragmatic criteria for ascribing these multi-leveled intentions as well. And this he explicitly does; regarding the moral stance, which is based upon these conditions, Dennett asserts: "the second choice (of

¹ All pagination in the text is from Daniel Dennett, *Brainstorms* (Montgomery, Vermont: Bradford Books, 1978).

moral commitment) is like the first [of ascribing intentionality] in being just a choice, relative to ends and desires and not provably right or wrong" (p. 241).²

Metaphysical personhood. At this point, however, things begin to get somewhat turbid. Although these conditions are necessary, Dennett feels that it is "hard to say whether they are jointly sufficient conditions for moral personhood" (p. 269). His worry seems to be based on the faint possibility that there may be a notion (presumably veridical) of "metaphysical personhood" which is distinct from moral personhood; and, *if* this is the case, he suggests that "there seems to be every reason to believe that metaphysical personhood is a necessary condition of moral personhood" (p. 269). This condition could perhaps be best expressed:

- (v) it really is a person (or, it possesses some metaphysically real property of personhood).

Dennett hints elsewhere (e.g., p. 285) that this condition would combine with (i) through (iv) to provide jointly sufficient conditions for moral commitment. However, it is hard to see why he would be tempted to think that (v) could be true or even relevant, given his pragmatism and lack of ontological commitment elsewhere, as well as his apparent doubts that (v) would be verifiable (see p. 285).

So, it is not surprising that Dennett's other remarks in this vein are more consistent with the position earlier outlined. The notion of metaphysical personhood is significantly diluted by his agreement that "our treating him or her or it in this certain [i.e., moral] way is somehow and to some extent *constitutive* of its being a person" (p. 270, emphasis mine). Along similar lines, he expressly describes a metaphysical person as "an entity to which states of consciousness or self-consciousness are *ascribed*" (p. 268, emphasis mine). In fact, he at one point almost hopefully remarks that "In the end we may come to realize that the concept of a person [presumably in the metaphysical sense] is incoherent and obsolete" (p. 267).

² However, there is for Dennett at least one being regarding which my stance is not a matter of choice, namely myself. Note his various claims: "I *must* view myself as a person" (p. 255, emphasis mine); "I cannot help but have a picture of myself as an intentional system" (p. 254); and "I am a person and so are you. That much is beyond doubt" (p. 267). Counting against this view, however, is his remark that under certain circumstances "we cannot even tell in our own cases if we are persons" (p. 285). This seems to suggest that we are able to choose to adopt a less than personal stance toward ourselves; but, given Dennett's view that only persons can make choices, this would be logically proscribed.

This, combined with Dennett's own readiness to adopt the moral stance towards those who satisfy conditions (i) through (iv), suggests that Dennett does expect us to take his pragmatic conditions as jointly sufficient for moral commitment, at least for everyday practical purposes. However, he is never willing to fully commit himself to their logical sufficiency.

2. Some Objections

Consistent application of Dennett's position results in the net of moral commitment being cast too narrowly in some cases and too widely in others. A few notable examples follow.

(a) *Dennett's Conservatism*

First of all, are there any beings which do not satisfy one or more of Dennett's necessary conditions, yet to which we nevertheless persist in attributing moral value?

Cases when behavior can be explained successfully without attributing the necessary conditions. Consider to begin with the most extreme case, in which even the primary condition of intentionality can be dispensed with. Surely there are persons (as the word is ordinarily used) whose behavior can be successfully explained and predicted from the physical or design stance, even as we continue to look at them from a moral point of view. Dennett frequently mentions the case of the mentally ill; often, he says, their behavior is much more successfully predicted by adopting the design — or even the physical — stance. But, surely, their rights as persons are not thereby lost; for we believe that their right to treatment and to being returned to full health is in no way inferior to the right of, say, the *physically* ill to proper treatment.

The case of small children is likewise problematic. The behavior of a tiny infant is clearly not easily predicted from the intentional stance. Yet our moral commitment to infants is immense; their rights to physical necessities, to love, and to preparation for dealing with life are not a matter of controversy (in fact, the common wisdom has it that their right to life supersedes all others when it comes to getting a place in the lifeboat!). Furthermore, if Dennett is right there is no way to account for the fact that we are not puzzled by the general granting of certain rights (such as proper prenatal care) to unborn humans.

Many other similar cases could be cited. The comatose, perhaps dying human may often be successfully dealt with without the assumption of intentionality; yet we feel bound to keep any promises made to him, and to provide him with a dignified death. Perhaps even the simpleminded,

inflexible individual who is wed to his routine and attitudes may not require our adoption of the intentional stance; but he does not thereby lose moral value.

Lest the abandonment of the assumption of rationality seems a rather contrived move in some of these cases, it should be noted that such extremes are not necessary for the short-circuiting of the moral stance. For even if one clings to the intentional stance, in these examples and many others one can successfully dispense with the ascription of second and third order intentions. This difficulty is compounded by Dennett's appeal to Lloyd Morgan's Canon of Parsimony in recommending that we "ascribe the simplest, least sophisticated, lowest order beliefs, desires, and so on, that will account for their behavior" (p. 274). But, whenever we find that we can account for a being's behavior with even two levels of intentionality, a necessary condition for personhood has not been satisfied (*viz.*, (iii)) and we are foreboded from adopting the moral stance. So, while things such as intentionality, reciprocity, communication, and consciousness may indeed have a bearing on moral commitment, it surely cannot be in the way that Dennett has suggested.

Cases when my goal is other than explaining and predicting behavior. To repeat, for Dennett the choice of stance is "relative to ends and desires," with the usual end being to explain and predict behavior. Now, if one's aim regarding a system is different, it is possible that one's stance will be different also. Dennett himself recognizes this possibility:

One can switch stances at will without involving oneself in any inconsistencies or inhumanities, adopting the intentional stance in one's role as opponent, the design stance in one's role as redesigner, and the physical stance in one's role as repairman. (p. 7)

Dennett refers here to a chess-playing computer, but he could as easily be talking about a human being; this would seem to suggest that any entity does not remain an intentional system by virtue of the ascription of intentionality under one set of circumstances, but may repeatedly lose and regain that status as one's goals shift. This is buttressed by his remark elsewhere that "The computer is an intentional system *in these instances* . . . just because it succumbs to a certain *stance* adopted toward it, namely the intentional stance. . . ." (p. 271, first emphasis mine).³ So, if

³ It is true that, contrary to these passages, he may at times be taken to suggest that once I adopt an intentional stance toward a being, it remains an intentional system relative to me even though I may alter that stance for some reason. For example: "An intentional system is a system whose behavior can be (*at least sometimes*) explained and predicted by . . . beliefs and desires . . ." (p. 271, emphasis mine). But to so interpret this passage would count against his clearer claims elsewhere, and would furthermore raise new problems (e.g., must I grant that my teddy bear is now an intentional system since it once

my goal is to make you look like a fool at a party, I may adopt the design stance and privately hypnotize you. If my goal is to keep you from inhibiting my robbery attempt, I may adopt the physical stance and tie you up. If my goal is to find some healthy kidneys for my ailing father, I may adopt the physical stance and cut yours out. In any such case, no talk of morality is permitted; for a necessary condition has not been met, due to the nature of my goal.

Similarly, in some cases I may find that my *only* goal toward you is a moral goal; I may have no interest in explaining, predicting, controlling, etc., your behavior. Rather, my only goal may be to behave toward you morally. This, however, is ruled out by Dennett, for such a choice has not been preceded by the prior explanatory choice.

One obvious rejoinder to such criticisms is that, in each of the examples mentioned, the explanation and prediction of behavior is still present as a goal, albeit a subordinate and perhaps subconscious one which likewise serves as a means toward the explicit goal; without such an intermediate goal, it might be argued, one would be hindered in achieving the larger goal. However, even if such a view were persuasive, it would not be satisfactory; for there is yet another category of beings who are indisputably persons yet whom are not accounted for by this revision.

Cases when I have no goal. In many cases, I may have no goal whatsoever regarding a "person". How can I be said to have any moral obligation to a hungry Biafran, to a future generation, or even to a child hit by a hit-and-run driver as I look on? Having no relationship to them, it may be that I have no goal toward them, and thus do not attribute intentionality, reciprocity, etc., to them.

It may be that Dennett would expand the moral community to include all those to whom I *would* ascribe intentionality, reciprocity, etc., *if* I had some goal toward them. But, again, as it stands Dennett's account of moral commitment includes no such provision.

The achievement of Dennett's aims for psychology. As Dennett sees it, the ultimate aim of psychology is as follows: "In the end, we want to be able to explain the intelligence of man, or beast, in terms of his design," having fallen short of this "whenever we stop in an explanation at the intentional level" (p. 12). Furthermore: "Eventually we end up, following this process, by prediction from the design stance; we end up, that is, dropping the assumption of rationality" (p. 10).

Clearly, once this aim is attained a necessary condition for the moral stance will have been jettisoned (at least for psychologists), thus making it a mistake to view any human as the proper object of moral commitment. But, perhaps this result would not be delayed until the aim is achieved. For

served — and for all I know may yet again serve — as my closest confidant?).

it may be that higher orders of intentionality would be dropped *before* the dropping of intentionality itself; and the moral stance would be eliminated with the elimination of the *first* multi-leveled intentionality (presumably consciousness).

Admittedly, the chances of this aim being reached are not very good. Even in the case of computers, whose design is known in detail, it is often necessary to adopt the intentional stance. But the point is that this is Dennett's avowed aim and it is, he thinks, in principle attainable. It must be noted, therefore, that achievement of his objective would, on his view, necessitate the abandonment of all moral commitment by those who are able to implement this mature psychology.

Before moving on to another sort of problem, one point about all the foregoing cases must be stressed. Not only would the circle of moral commitment be drawn too tightly, but it could conceivably be drawn *differently* by every person (and justifiably so). So, you may adopt the moral stance toward me even as I correctly withhold all moral commitment toward you. Furthermore, the membership roster of the moral community could vary not only from person to person, but for a single person from time to time — always depending upon his goals and his current explanatory skills. It hardly needs to be remarked that any position which generates such a result must be seriously mistaken.

(b) Dennett's Liberalism

What of cases in which Dennett's schema suggests a moral stance toward beings which we normally do *not* include in the moral community? It will be helpful here to keep in mind that Dennett has described condition (iii) (communication) as sufficient for (iv) (consciousness); so, since they are also sufficient for (i) and (ii), the presence of either (iii) or (iv) will indicate that all of his necessary (and more-or-less sufficient) conditions have been satisfied.

Cases animal, vegetable. . . . The capacity to communicate is not infrequently ascribed to higher mammals such as apes and porpoises, even by some specialists in such matters. Dennett even grants that plants can be at least intentional systems (p. 272), and surely our animistic ancestors are not the only ones who have taken the further step of granting them higher orders of intentionality. And, to borrow Rorty's remark on this matter, "a similar effort of imagination will produce the same results for any chunk of the universe."⁴ That is to say, given the pragmatic point of view one could treat anything as if it were communicating. Yet we do not expect

⁴ Richard Rorty, "Functionalism, Machines, and Incorrigeability," *Journal of Philosophy* 69 (1972): 207.

those who so exercise their imagination to afford said chunk of the universe the same moral status as they afford, say, human beings. Dennett, however, has offered no means of making this moral differentiation. He would no doubt appeal here to the principle of parsimony. But the inappropriateness of employing Lloyd Morgan's Canon has already been noted: not only does it narrow the moral circle too much by directing that we strive to abandon multiple levels of intentionality wherever possible; but it is also applied arbitrarily, with the basic assumption of intentionality (condition (i)) mysteriously exempt from it (see p. 274). To these considerations another might be added: people simply are not very parsimonious in granting higher levels of intentionality, yet still they morally differentiate.

. . . and even mineral. Even if higher order intentions are subjected to the razor, the problem of liberality is still not fully resolved. Dennett can "easily contemplate the existence of biologically very different persons" (p. 267); but is he ready to accept computers as persons? It may seem quite natural to someone to attribute third order intentions to, say, a very good chess-playing machine; for achieving my goal of winning may be facilitated by assuming that it *intends* that I *believe* that it *plans* to follow a certain strategy, when in fact it may be misleading me. Even Dennett acknowledges that it is possible for there to be a computer which communicates (i.e., satisfies condition (iii)) to even the most parsimonious among us (pp. 280, 281, n.). It would seem that, in the presence of such a computer, he would have no choice but to grant it complete moral personhood.⁵

If there are only a few with whose intuitions this is consonant, there are even fewer who would endorse a further possible consequence. If Dennett's aim for psychology is in fact realized and we are someday able to explain humans entirely from the design stance, it is likely that it will be done largely with the help of computers. Suppose I design the computer which accomplishes that task. Let it also be a learning computer, one so sophisticated that I cannot possibly relate to it without taking it to be a reciprocator, communicator, etc. Now, if Dennett is right, so far as *I* am concerned both the computer and I are moral agents; but I must at the same time grant that from the computer's point of view, it (he/she?) would be *mistaken* to view *me* as a moral agent, for the computer explains and predicts my behavior from the design point of view. So, even as I grant it

⁵ One philosopher, of course, who is quite open to this position is Hilary Putnam. See "Robots: Machines or Artificially Created Life?" in Hilary Putnam, *Mind, Language, and Reality* (Cambridge: Cambridge University Press, 1975), p. 407.

rights, I must permit it to deny them of me. Can an account of moral commitment which leads to such conclusions be correct?

3. Conclusion

And so, Dennett's conditions for moral commitment may clearly be faulted as being both too strong and too weak. How exactly has he gone wrong? It would not be entirely on the mark to say that he fails because he makes moral commitment dependent upon decisions which are largely irrelevant to morality. For Rawls makes such a move with some measure of success, by having the principles of justice selected by persons acting strictly from rational self-interest. An important difference, however, is that this non-moral consideration has bearing for Rawls only in the theoretical original position; in everyday life, rational self-interest is to be displaced by adherence to the principles selected. Dennett, however, would have our moral attitudes in a constant state of flux; whether or not any given object has moral personhood would depend upon each person's individual goals and skills at offering explanations with minimal intentionality. As such, the decision tells us more about the decision maker than it does about the being in question. Until this flaw is corrected, Dennett's position must remain inadequate.⁶

⁶ I am grateful to Tyler Burge for valuable criticisms of an earlier version of this paper.