

THE ROLE OF HOMUNCULI IN PSYCHOLOGY

ANDREW WARD

In recent years the view that a significant portion of intentional human behavior is *caused* by, and hence explainable and predictable in terms of, mental states which are representational in character has come in for a number of criticisms. One of the most prevalent forms of criticism can be culled from the writings of Daniel Dennett and turns on his claim that "nothing is intrinsically a representation of anything; something is a representation only for or to someone" (AAIP, p. 122).¹ Put differently, Dennett's claim is that "any representation or system of representations requires at least one *user* of the system who is external to the system" (CCC, p. 101).² This means that if a proponent of the view that intentional behavior "has proven to be of such a nature that the only satisfactory theories will be those in which [...] representations will play a role" (BWMR, p. 41), often called representationism, accounts for such behavior in terms of specific relations between the organism and a system of representations, then there must be some user of the representations who is external to the system—an "exempt agent". However, since the exempt agent "must be capable of comprehension, and [...] have beliefs and goals (so it can *use* the representation[s] to inform itself

¹In what follows AIPP refers to "Artificial Intelligence as Philosophy and as Psychology", BWMR to "Brain Writing and Mind Reading", CCC to "A Cure for the Common Code", IS to "Intentional Systems", SS to "Skinner Skinned", and WLE to "Why the Law of Effect Will Not Go Away", all in Daniel Dennett's *Brainstorms* (Cambridge, MA: MIT Press, 1981); SMR refers to "Styles of Mental Representation," *Proceedings of the Aristotelian Society*, n.s., v. 83, pp. 213-226 (1982/83); TB to "True Believers: The Intentional Strategy and Why it Works," in *Scientific Explanation* edited by A.F. Heath (Oxford: Clarendon Press, 1981), pp. 53-75; TK to "Three Kinds of Intentional Psychology," in *Reduction, Time and Reality*, edited by R. Healey (Cambridge: Cambridge University Press, 1981), pp. 37-61; and CC to *Content and Consciousness* (London: Routledge and Kegan Paul, 1986).

²Marx W. Wartofsky, "Epistemology Historicized," in *Naturalistic Epistemology: A Symposium of Two Decades*, edited by Abner Shimony and Debra Nails (Dordrecht: D. Reidel Publishing Co., 1987), p. 36, puts the point this way: "Nothing is a representation as such. It comes to be a representation by being *taken* as a representation, or *made* as a representation."

and thus assist it in achieving its goals)" (AIPP, p. 122), it follows that it too will exhibit intentional behavior. Now either we can follow through on the representationalist programme and account for this behavior in terms of relations to representations, or we can opt for some non-representationalist account of the exempt agent's intentional behavior. In the first case we are led by the principle that "any representation or system of representations requires at least one *user* of the system who is external to the system" to posit yet another exempt agent for whom the same question arises, viz., how are we to account for its intentional behavior? Moreover, by consistently following through on the representationalist programme we seem led to ask the question at each stage of analysis and so posit an infinite regress of exempt agents. (see AIPP, pp. 122-123; also see CC, p. 87) But since "a non-question-begging psychology will be a psychology that makes no ultimate appeal to unexplained intelligence" (WLE, p. 83; also see SS, p. 61), and because exempt agents exhibiting intentional behavior are agents that exhibit unexplained intelligence, Dennett concludes that a non-question-begging representationalist psychology is impossible (see SMR, pp. 213 ff).

Turning to the second case, it is difficult to see why an advocate of this resolution to Dennett's criticism would grant the representationalist's initial supposition that intentional behavior "has proven to be of such a nature that the only satisfactory theories will be those in which [...] representations will play a role." (BWMR, p. 41) After all, if an account of the intentional behavior of exempt agents is forthcoming without an appeal to representations, then why would not the same sort of explanation obviate any need for representations in the case of the system from which the exempt agents are exempt? It would seem that the role of representations on such an account is explanatorily superfluous. Thus, to opt for a non-representationalist account of the intentional behaviors of exempt agents seems tantamount to a rejection of representationalism. The upshot is that in either case Dennett's remarks seem telling against a representationalist account of intentional behavior.

Viewing the problem for representationalist psychology in terms of the above dilemma, the ideal move for the representationalist would seem to be to grasp the dilemma by its first horn and provide an account that does not fall prey to a vicious infinite regress. But is such an account forthcoming? Here it is important to see that the crux of Dennett's argument is that the use exempt agents make of representations is intentional. It is this fact which requires positing an infinite regress of exempt agents. Accordingly, there are at least two different ways for the representationalist to grasp the first horn of the dilemma. The first, Dennett's own way, is to begin at the molar level with an intentionally characterized system, decompose it into "an organization of

subsystems [homunculi], each of which could be viewed as an intentional system" (AIPP, p. 123) and then continue to break down these subsystems still further until all intentionally characterized subsystems have been discharged in favor of "homunculi so stupid [...] that they can be, as one says, 'replaced by a machine'." (AIPP, p. 124) Put differently, the idea is for the representationalist to progressively decompose the original intentional system into teams of "smaller, individually less talented and more specialized homunculi"³ until finally reaching the point where the behaviors of the homunculi require only "problem or task descriptions that are obviously mechanistic" (WLE, pp. 80, 81). At this final stage of analysis the homunculi do not exhibit any intentional behaviors at all, and so are not agents, exempt or otherwise (see SMR, pp. 220-221).

In contrast, a second way for the representationalist to grasp the dilemma by its first horn is to say that the *use* exempt agents make of representations is *not* intentional. In this case there is no need to posit any further exempt agents in order to account for the original system's intentional behavior, and so the regress never starts. The question for this way of grasping the first horn of the dilemma is whether, in accounting for an organism's intentional behavior, sense can be made of the claim that the account must make reference to representations that are used in a non-intentional manner by exempt agents.

Given these two alternatives, which one ought the representationalist to opt for? Beginning with Dennett's proposed solution, one objection it immediately faces is that its adoption seems tantamount to giving up on representationalism. Dennett's account, the objection runs, solves the problem of the infinite regress of exempt agents by terminating the process of decompositional analysis with an organization of "elements familiar to the biologists" (AIPP, p. 110). At this final level of analysis intentional descriptions give way to non-intentional, biological descriptions. The problem is that this amounts to the claim that all intentional descriptions are ultimately decomposable into a complex set of non-intentional descriptions (see WLE, p. 81).⁴ As William Bechtel puts it, for Dennett "subpersonal analyses are to explain how a mechanism could perform *as* something

³William G. Lycan, "Form, Function and Feel", *The Journal of Philosophy*, 78, p. 28 (1981). Also see Philip Cam, "Dennett on Intelligent Storage," *Philosophy and Phenomenological Research*, 45, p. 257 (1984), and Joseph Margolis, "Conceptual and Methodological Links between Cognitive Psychology and Philosophy of Psychology," *Cognition and Brain Theory*, 5, pp. 316-317 (1982).

⁴See Joseph Margolis, "Conceptual and Methodological Links between Cognitive Psychology and Philosophy of Psychology," p. 317, and Joseph Margolis, "Psychology and Its Methodological Options," in *Psychology; Designing the Discipline*, edited by Joseph Margolis (Oxford: Basil Blackwell, 1986), p. 41.

intentional.”⁵ In this case though, it follows that intentional descriptions involving representations are, in principle, eliminable.⁶ Moreover, since beliefs, desires and other propositional attitudes are paradigm instances of intentional states, it seems to follow that one ought to resist making ascriptions of propositional attitudes, at least as conceived by traditional folk psychology, to organisms.

Now it is true that, contrary to this line of objection, Dennett says that “people really do have beliefs and desires [...] just the way they really have centers of gravity and the earth has an equator” (TK, p. 46). But what exactly does he mean by this? Using terms introduced by Hans Reichenbach, Dennett suggests that the beliefs and desires of a scientific psychology are *abstracta*, where *abstracta* are “calculation bound entities” (TK, p. 46). It is in this respect, says Dennett, that beliefs and desires are just as “real” as the earth’s equator or centers of gravity, both of which are also *abstracta* (TK, p. 46). As it stands though, this claim can be fleshed out in either of two ways. On the one hand it can be read as saying that beliefs and desires are entities having only instrumental existence, while on the other it can be read as an account which is, in Jerry Fodor’s terms, realist.⁷ When inclined towards an instrumentalist reading, Dennett says that “intentional theory is vacuous as psychology because it presupposes and does not explain rationality or intelligence [and that, as a result] Skinner is right in recognizing that intentionality can be no *foundation* for psychology, and right also to look for purely mechanistic regularities in the activities of [...] internal system whose design approaches the optimal (relative to some ends)” (IS, p. 15). The problem for the representationalist is that such an approach construes intentional behavior as essentially epiphenomenal. That is to say, the instrumentalist who embraces Skinner’s rejection of intentionality as a foundation for (scientific) psychology is committed to saying that *all* behaviors of homunculi are to be ultimately accounted for by the behavior of subsystems of non-intelligent (presumably biological) structures. The consequence is that while “representational talk” may sometimes be useful for predicting behavior, organisms do not *really* exhibit intentional behavior (see IS, p. 7 and SS, pp. 61-62). As Dennett says, “If we are to have any adequate *analysis* of creativity, invention, intelligence, it must be one in

⁵William Bechtel, “Realism, Instrumentalism and the Intentional Stance,” *Cognitive Science*, 9, p. 475 (1985).

⁶See Robert C. Richardson, “Intentional Realism or Intentional Instrumentalism,” *Cognition and Brain Theory*, 3, pp. 129, 130 (1980).

⁷See Jerry Fodor, “Fodor’s Guide to Mental Representation: The Intelligent Auntie’s Vade-Mecum” *Mind*, 94, pp. 78ff. (1985). Also see Joseph Margolis, “The Trouble with Hamunculus Theories,” *Philosophy of Science*, 47: 244-259 (1980).

which intelligence is analyzed into something none of whose parts is intelligence" (WLE, p. 89). The upshot is that the (ultimately) non-intentional character of sub-personal (homunculi) psychology, together with the need to avoid a question-begging explanatory regress, infects what Dennett calls "pure intentional system theory" (TK, p. 50) so that representations are, *at best*, convenient fictions. (see IS, p. 7).⁸

In contrast, on a realist reading the beliefs and desires of a scientific psychology are legitimate abstract entities that are neither reducible to nor eliminable in favor of non-intentional states. As Joseph Margolis puts it, on a realist reading "animals and humans may be ascribed [at the molar level] actual mental states such as believing, [...] desiring, [...] and the like."⁹ Perhaps these are not the beliefs and desires that people ordinarily talk about, but that does not compromise a realist construal of the *abstracta*. Rather, the *abstracta* are rigorously defined concepts to which reference *must* be made in order to adequately account for a system's intentional behavior (see TB, pp. 64ff.).

However, a problem begins when Dennett attempts to connect these scientifically rigorous entities to the progressive decompositions of organizations of homunculi required in order to avoid the infinite regress problem. At each stage of decomposition the question may be asked: do the homunculi exhibit intentional behaviors or not? If they do, then to avoid a "question-begging psychology" the process of decomposition must continue and each subsystem of intentionally characterized homunculi must be analyzed into yet another subsystem of individually less talented, more specialized homunculi of which the same question must then be asked. In order to avoid an infinite regress there must come a point when the question is answered negatively. (see TB, pp. 72-73). In this case though, at the final stage of analysis the representationalist seems committed to an account of intentional behavior given wholly in terms of the organization of non-intentional "elements familiar to the biologists." (AIPP, p. 110). As Philip Cam puts it:

⁸See Lilly-Marlene Russow, "Dennett, Mental Images, and Images in Context," *Philosophy and Phenomenological Research*, 45, p. 582 (1985). Also see Joseph Margolis, "Psychology and Its Methodological Options," p. 40. In "Intentional Systems in Cognitive Ethology: The 'Panglossian Paradigm' Defended," *The Behavioral and Brain Sciences*, 6, p. 380 (1983), Dennett distinguishes fictionalist instrumentalism from non-fictionalist instrumentalism. Whereas his view in "Intentional Systems" was, presumably, fictionalist, his view in this later paper is non-fictionalist (see also TB, p. 67). In particular, Dennett says that "attributions of belief and desire are not just 'convenient fictions'; there are plenty of honest-to-goodness instrumentalist *truth*..." (p. 380). In what follows, I will treat such a view as realist (Dennett calls it 'mild realism' in TB, p. 67).

⁹Joseph Margolis, "The Trouble with Homunculus Theories," p. 246. Also see Jerry Fodor, "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum," p. 78.

By a series of such moves these "homunculi" are broken down into others that at each stage explain the functioning of the systems posited on the next highest level. Eventually intentional characterizations give way to *purely extensional ones*.¹⁰

The point to notice through, is that even if this sort of procedures can be carried through, the representationalist has slipped from the first horn of Dennett's dilemma to the second. In particular, an account of intentional activity has been proposed such that for each intentional state "there will be a functionally salient internal state of the machinery" that *will account for* the intentional state (TB, p. 72). Here, it is not just that intentional systems are either eliminable in favor of or are realized by physical systems, it is that intentional systems are either eliminable in favor of or *reducible to* physical systems. Hence, the cost of reading Dennett along realist lines seems to be an abandonment of representationalism. The upshot is that regardless of whether one gives an instrumentalist or realist reading to Dennett's solution of the infinite regress problem, the cost would seem to be a rejection of representationalism.

So what of the other way to avoid the regress problem? Will this allow the representationalist to avoid the problems of Dennett's account while retaining some genuine form of representationalism? As noted above, the second alternative is viable only if sense can be made of exempt agents using representations non-intentionally. Along these lines, it is useful to recall the distinction drawn by W. V. O. Quine between behavior which *fits* a rule and behavior which is *guided* by a rule. In particular:

Behavior *fits* a rule whenever it conforms to it; whenever the rule truly describes the behavior. But the behavior is not *guided* by the rule unless the behavior knows the rule and can state it.¹¹

¹⁰Philip Cam, "Dennett on Intelligent Storage," p. 257 (my emphasis). Also see Robert N. McCauley, "The Role of Cognitive Explanations in Psychology," *Behaviorism*, 15, pp. 29ff. (1987). As McCauley says, "to explain the organism's intelligence by the intelligence of its putative cognitive sub-systems still leaves us within the realm of intelligence and only postpones the inevitable [...] An unending regress of refinancing one intelligence loan with others is *no solution* to the problem of explaining human behavior" (p. 30, my emphasis).

¹¹W. V. O. Quine, "Methodological Reflections on Current Linguistic Theory," in *Semantics of Natural Language*, second edition, edited by Donald Davidson and Gilbert Harman (Dordrecht: D. Reidel Publishing Company, 1972), p. 442. Also see Wilfrid Sellars, "Empiricism and the Philosophy of Mind," in *Science, Perception and Reality* (London: Routledge and Kegan Paul, 1963), pp. 166ff.

Accordingly, to say that an organism's behavior is guided by a rule requires attributing both agency and knowledge¹² to the organism, whereas to say that an organism's behavior fits a rule requires, at most, only that any adequate account of organism's behavior make reference to some rule. In the present context the parallel is that if representations are used by exempt agents in the sense that they guide the behavior of exempt agents, then Dennett's infinite regress problem must be dealt with. After all, just as behavior that is guided by a rule presupposes an attribution of agency, so too behavior that is guided by representations presupposes comprehension of the representations and the desire to attain certain goals, both of which are intentional.¹³ It follows that nothing would be gained for the representationalist to say that exempt agents use representations in a manner analogous to how some behaviors are rule-guided. On such an account the representationalist would still be left with the problem of explaining the intentional behavior of exempt agents, and so too with a question-begging psychology. The upshot is that exempt agents can be said to use representations non-intentionally only if their behavior is *not* guided by those representations.

The failure of modelling the use of representations by exempt agents on behavior that is guided by a rule leads to the second part of Quine's distinction. In particular, using the notion of behavior that fits a rule as a model, is it enough for the representationalist to say that the behaviors of exempt agents fit a rule? Here it is useful to recall what motivated the representationalist to accept talk about exempt agents in the first place, viz., Dennett's claim that a representation could be a representation "only for or to someone." What lies behind this claim is the notion that if representations are to serve any function they must enter into the etiology of an organism's molar behavior and that they do this by entering into the etiology of the relevant exempt agents' behavior.¹⁴ But in what sense does the representationalist want to say that representations enter into such behavior? Certainly one way for them to do so is for exempt agents to have intentional relations to the relevant representations. As noted above though,

¹²Though the knowledge may be tacit; see SMR, pp. 214ff.

¹³By comprehension I do not mean awareness. I am not saying that the subject whose behavior is guided by the rules is *aware* of the rules, only that whatever else such behavior is, it is intentional. See Jerry Fodor, "The Appeal to Tacit Knowledge in Psychological Explanations," in *Representations* (Cambridge, MA: MIT Press, 1981), pp. 63-78. Also see Michael D. Root, "Language, Rules and Complex Behavior," in *Minnesota Studies in the Philosophy of Science VII: Language, Mind, and Knowledge*, edited by Keith Gunderson (Minneapolis: University of Minnesota Press, 1975), pp. 321-343.

¹⁴See Jerry Fodor, "Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum," p. 78; also see Fodor's "Introduction: Something on the State of the Art," in *Representations*, pp. 1-31.

this leads the representationalist to the use of a non-analyzed intentional characterization and so to a question-begging psychology. Therefore, suppose that the representationalist says the representations enter into the etiology of the molar behavior of organisms in that the behaviors of the relevant exempt agents *fit* rules that make reference to these representations. Here the idea is that the representationalist is committed to saying that, as a matter of empirical fact, any adequate explanation of an organism's molar intentional behavior must appeal to a description of certain rule-fitting behaviors by exempt agents (see TB, p. 64). Put differently, the representationalist says that the best theory of the molar intentional behaviors of organisms is one that requires that reference be made to those rules governing the interactions of representations.

On the account being suggested, the representationalist concedes to Dennett that "nothing is intrinsically a representation of anything; something is a representation only for or to someone" (AIPP, p. 122) by granting the presence of exempt agents for whom representations *are* representations. Moreover, the representationalist even concedes to Dennett the possibility that some (though not all) relations between exempt agents and representations are intentional. Indeed, my own view is that it *will* be necessary to attribute intentional behaviors to exempt agents in order to account for self-ascriptions of identity by organisms.¹⁵ Still, whether or not I am correct about this matter, what is important is that the representationalist cannot and does *not* grant to Dennett that accounting for the intentional behavior of organisms at the molar level requires making reference to intentional behaviors of exempt agents. Rather, the representationalist says two things. First, it is a fact, subject to empirical verification or falsification, that any adequate account of the relevant behavior of exempt agents must use descriptions that make reference to representations related to one another in determinate ways. Here, following the lead of Noam Chomsky, the point is that there is nothing more involved in saying that representations enter into the etiology of the behavior of exempt agents than that these representations, and their relations, are constituent elements of the explanatory theory that best accounts for the intentional behavior of organisms.¹⁶ On such an account exempt agents do not guide their behaviors according to rules linking representations to one another because exempt agents do *not* interpret rules. Rather, as Wittgenstein says:

¹⁵Roughly, what I have in mind here is a Humean account of (self)consciousness. For remarks in this direction see my "Hume, Demonstratives and Self-Ascriptions of Identity," *Hume Studies*, 11: 69-93 (1985).

¹⁶See Noam Chomsky, *Knowledge of Language: Its Nature, Origin, and Use* (New York: Praeger Publishers, 1986), pp. 249-253, 260f.

there is a way of grasping a rule which is *not* an interpretation, but which is exhibited in what we call "obeying the rule" and "going against it" in actual cases.¹⁷

To say that exempt agents "obey a rule" is to say that there is an observed regularity in the organisms' molar behavior which is best accounted for by making reference to exempt agents whose behaviors fit rules governing the interactions of representations. (see TB, pp. 64, 70). The prior question of whether a given organism exhibits the requisite regularity in its behavior to warrant attributing intentionality to its behavior is something that can be decided only relative to the organism's role within a community in which attributions of intentional behaviors are made. It is this role that distinguishes the sense in which an organism follows a rule that leads to an attribution of intentionality from the sense in which a lightning-thunder sequence follows a rule not leading to an attribution of intentionality.¹⁸ The second thing the representationalist says is that descriptions of the behavior of exempt agents making reference to representations are neither reducible to nor eliminable in favor of descriptions that do not make reference to such representations.¹⁹

Where then does this leave the representationalist? The answer, I think, is that it leaves the representationalist with a defensible account of the sense in which a significant portion of human behavior "is caused by, and hence explainable and predictable in terms of mental states which need to be intentionally characterized."²⁰ To vary what Dennett says in support of the intentional stance to suit the present point, there are patterns in human behavior that are describable in representationalist terms, and *only* in those terms, and which support important generalizations and predictions (see TB, p. 64). It is in this sense that the representationalist may rightly insist that (intentional) human behavior "has proven to be of such a nature that the only satisfactory theories will be those in which [...] *representations* play a role" (BWMR, p. 41). By allowing that the use exempt agents make of representations is not (always) intentional, the representationalist is able to defuse Dennett's regress by grasping the first horn of the dilemma.

St. Olaf College

¹⁷Ludwig Wittgenstein, *Philosophical Investigations* (3rd edition), translated by G.E.M. Anscombe (New York: Macmillan Publishing Co., 1979).

¹⁸See Wilfred Sellars, "Empiricism and the Philosophy of Mind," pp. 166ff.

¹⁹It is important to notice that nothing has been said that would preclude the compatibility of representationalism with the view that every psychological event is a neurological event (token physicalism).

²⁰Paul Yu and Gary Fuller, "A Critique of Dennett," *Synthese*, 66, p. 457 (1986).