THE COMPATIBILITY OF PSYCHOLOGICAL NATURALISM AND REPRESENTATIONALISM

Andrew Ward
Georgia Institute of Technology

Fred Dretske begins chapter three of his *Explaining Behavior* with the following observations:

> Some behavior is the expression of intelligent thought and purpose. Clyde goes to the kitchen because he wants another beer and thinks there is one left in the refrigerator. Whether or not they are *causes* of behavior, Clyde's reasons — his desire for a beer and his belief that there is one in the fridge — are certainly thought to *explain* his behavior. They tell us *why* he made the trip to the kitchen.[1]

There are at least two important ideas in Dretske's story of Clyde. First, not all explanations are narrowly causal explanations. As Wittgenstein is reported to have said, sometimes an explanation consists in "the understanding which consists in seeing connections…"[2] Second, at least some of our behaviors require an ability to represent the world. As John Haugeland writes:

> A sophisticated system (organism) designed (evolved) to maximize some end (such as survival) must in general adjust its behavior to specific features, structures, or configurations of its environment in ways that could not have been fully prearranged in its design. If the relevant features are reliably present and manifest to the system (via some signal) whenever the adjustments must be made, then they need not be represented. Thus, plants that track the sun with their leaves needn't represent it or its position, because the tracking can be guided directly by the sun itself. But if the relevant features are not always present (manifest), then they can, at least in some cases, be represented; that is, something else can stand in for them, with the power to guide behavior in their stead. That which stands in for something else in this way is a *representation*…[3]

---

[1] Dretske, p. 51.
[2] Quoted Monk, p. 451.
[3] Haugeland, "Representational Genera," p. 172.

Relative to Dretske's example, perhaps if Clyde's kitchen never changed, if his pathway to the refrigerator was always unobstructed, and if there was always beer in the refrigerator, then his behavior would not need a representational explanation. Suffice to say though, this would be an unusual situation. Thus, Clyde's behavior seems a likely candidate for having a representational explanation. Moreover, the complexity exemplified in Clyde's behavior is typical of many behaviors that human beings exhibit.[4] Accordingly, the assumption that some significant portion of human behaviors requires an ability to represent the world appears quite commonsensical. This has led many people to contend that the only plausible psychology that could succeed in explaining a significant portion of human behaviors must make use of mental representations. Commonsense notwithstanding, a central issue for theories making use of mental representations concerns their compatibility with our "emerging neuroscientific picture of living organisms."[5] How, in the words of Stephen Stich, can representational properties of mental states "be explained in terms that are compatible with the broader, physicalistic view of nature provided by the natural sciences"?[6] In what follows I will outline a strategy of integrating mental representations into the natural world.

A good place to start is with why someone would think that the compatibility of mental representations with an "emerging neuroscientific picture of living organisms" is problematic. There are at least three kinds of answers. First, there is the question of how something like a neurological state can represent anything at all. Paradigmatic instances of representations involve some sort of similitude relationship.[7] For instance, a photograph is representative of what has been photographed because it bears a certain resemblance to what has been photographed. The problem is that using this as a paradigm of representation makes it difficult to understand how a neurological state could represent anything other than, at best, some other neurological state.[8] Second, there is the question of what role mental representations play in causing behavior. For instance, in the case of Clyde, suppose that we were to totally separate the kind of explanations in which mental representations play a role

_____

[4] Though I restrict attention to human beings in this paper, there is no reason that the account offered could not be extended beyond the case of human beings.

[5] Dretske, p. 51.

[6] Stich, p. 246. Also see Tye, p. 422.

[7] As Cummins, 1989: pp. 2-6, notes, this is the conception of representation that is common to both representations as "inFORMED mind-stuff" (e.g., Aristotle) and the later developments of the British empiricists (e.g., Berkeley) for whom representations were "images".

[8] As Cummins, 1989: p. 6, writes, "[L]ike symbols, neurophysiological states cannot represent things in virtue of resembling them. Advocates of symbols or neurophysiological states must ground representation in something other than similarity."

from causal explanations. In such a case, to the extent that the "emerging neuroscientific picture of living organisms" seeks for causal explanations, explanations involving mental representations will be at best superfluous, and at worst misleading. Third, there is the fact that representations are relational (observer-relative) in character. As Daniel Dennett writes, "nothing is intrinsically a representation of anything; something is a representation only for or to someone."[9] An example from Hilary Putnam neatly illustrates this character of representations:

> An ant is crawling on a patch of sand. As it crawls, it traces a line in the sand. By pure chance the line that it traces curves and recrosses itself in such a way that it ends up looking like a recognizable caricature of Winston Churchill. Has the ant traced a picture of Winston Churchill, a picture that *depicts* Churchill?

Most people would say, on a little reflection, that it has not. The ant, after all, has never seen Churchill, or even a picture of Churchill, and it had no intention of depicting Churchill. It simply traced a line (and even *that* was unintentional), a line that *we* can 'see as' a picture of Churchill.[10]

The problem is that mental representations were supposed to explain, amongst other things, intentional behavior. Clyde's behavior was described using representations because we were trying to account for his intentional behavior of seeking a beer in the refrigerator. If, like the plant tracking the sun across the sky, his behavior hadn't been intentional, there wouldn't be a need for a representational explanation. However, if it is Clyde who is, in some sense, making *use* of mental representations, then it seems that there must be a primitive, non-representational kind of intentionality in order for mental representation explanations to get off the ground. This appears to introduce a non-physical, mental element into explanations of a significant portion of

---

[9] Dennett, 1981: p. 122. Also see Dennett, 1981: p. 101, and Putnam, 1981: chapter 1. In contrast with some representationalists who claim that the formal, structural characteristics of representations are not "observer relative", Searle, 1990: p. 35, claims that the formal (syntactical/structural) character of representations is, like the non-formal, representational (semantic) character of representations, observer relative. What this means, writes Searle, 1990: p. 29, is that "[W]ithout a homunculus that stands outside the recursive decomposition, we do not even have a syntax to operate with." If correct, then the problems for representationalism posed by the relativity of representations cannot be resolved by considering only the formal (syntactic) characteristics of representations.

[10] Putnam, 1981: p. 1. Also see Haugeland, "The Intentionality All-Stars," pp. 129-130, who writes that "if some cyanide molecules in the Crab Nebula should happen to trace out the shape of the letters 'Hank Aaron', they would not name the homer Hammer (or even be the letters 'Hank Aaron')."

human behaviors. Although these problems are often treated separately, it is the last problem that is central.

To understand why the third problem is central, an important first step is to recognize that the relational character of mental representations seems to require positing constituent psychologically characterized subsystems *for whom* representations are representations. That is, if Dennett is right that "nothing is intrinsically a representation of anything; something is a representation only for or to someone", then attributing mental representations to a person is really a two-stage process. First, there is the stage in which we are led on the basis of observable behaviors (within a certain context) to attribute mental representations to the organism manifesting the behaviors. Second, there is the stage in which, if the organism in question is making use of the mental representations, we must attribute some "psychologically internal interpreter" of the mental representations. Such psychologically internal subsystems whose role is to interpret and so make use of mental representations are often called 'homunculi'. The point is that because representations have non-formal, representational properties only in virtue of their being representations "for or to someone", it follows that mental representations can enter into explanations of behaviors only in virtue of their *interpretation* by homunculi.[11] Thus, the representational behaviors of human beings are explained in terms of the *uses* homunculi make of *interpreted* mental representations.[12] An imaginative example from Jerry Fodor provides a good illustration of the idea:

> This is the way we tie our shoes: There is a little man who lives in one's head. The little man keeps a library. When one acts upon the intention to tie one's shoes, the little man fetches down a volume entitled *Tying One's Shoes*. The volume says such things as: "Take the left free end of the shoelace in the left hand. Cross the left free end of the shoelace over the right free end of the shoelace…" etc. … When the little man reads "take the left free end of the shoelace in the left hand," we imagine him ringing up the shop foreman in charge of grasping shoelaces. The shop foreman goes about supervising that activity in a way that is, in essence, a microcosm of tying one's shoe. Indeed, the shop foreman might be imagined to superintend a detail of wage slaves, whose functions include: searching representations of visual inputs for traces of shoelace, dispatching orders to flex and contract fingers on the left hand, etc.[13]

----

[11] Lycan, 1988: p. 5. As Dennett, 1981: p. 124, writes, "each homunculus has representations that it uses to execute its functions …"

[12] Searle, 1990: p. 28, calls this the semantic form of the homuncular fallacy.

[13] Fodor, 1987: pp. 23 - 24.

Continuing on with talk of homunculi, since the use homunculi make of mental representations is a function of their interpreting those representations, homunculi must "be capable of comprehension, and … have beliefs" so that they can use the representations to achieve their goals.[14] But now recall that the "emerging neuroscientific picture of living organisms" understands mental representations broadly as neurological states.[15] Thus, if we think of homunculi (which are themselves neural structures) as interpreters of neurological states or sets of neural states, then we have an answer to how neurological states can be representational. Neurological states have representational characteristics in a way analogous to that in which written or spoken words have representational characteristics. So understood, the representational character of neurological states is not more mysterious than that of spoken and written language. Indeed, divorced of its sentential commitments, this is pretty much the way that Francis Crick describes the way that distributed parts of the brain monitor how other parts of the brain — sets of neurons — interact with one another.[16] Moreover, we also have an answer as to how mental representations enter into explanations of an organism's behavior. Mental representations qua neural states or sets of states are interpreted by psychologically internal subsystems — homunculi qua neural structures — so as to achieve certain goals.[17] Again, the analogy can be drawn to the ways in which a story can enter into an explanation for why a person does what she or he does. The person reads the story and in the process of reading, interprets the information in such a way that certain behaviors are discovered to be consistent with the realization of desired goals, while others are inconsistent with their realization. Similarly, homunculi "read"/interpret mental representations in such a way that certain behaviors are discovered to be consistent with the realization of desired goals, while others are inconsistent with their realization. Fodor's "little man who lives in the head" is, in reality, a part of the head — a semi-autonomous neural module whose function is to interpret the representational information supplied by other parts of the brain.

––––––––––

[14] Dennett, 1981: p. 122. Also see Cummins, 1983: pp. 91-92.

[15] There are really two different views that can be teased out here. On the one hand, if one focuses on the symbolic character of mental representations, then the fact that they are instantiated by neural structures is not crucial. Anything that could instantiate the requisite symbolic structure would suffice — e.g., a computer. On the other hand, if one views mental representations as biological phenomena essentially, then, as Cummins, 1989: p. 6, writes, "[M]ental representations cannot be realized in, say, a digital computer, no matter how :brain-like" its architecture happens to be at some nonbiological level of description." For the purposes of this paper, it is not important to consider separately the two views.

[16] See, for example, Crick, pp. 203ff.

[17] See Haugeland, 2000: pp. 117ff.

Here though, we reach the crux of the problem. We have answered the three problems associated with mental representations by treating them on a par with ordinary spoken or written language. In effect, we have treated mental representations as if they were elements in a language of thought.[18] The problem is that this explanation seems really to be no explanation at all. We have explained mental representations by positing homunculi who themselves have a variety of rather complex psychological states, and so have answered the problems associated with mental representations by treating their representational characteristics as derivative — derivative of the psychology of interpreting homunculi. Just as users of English give the representational character of 'cat' to it, analogously the representational character of a neural state is given to it by the interpreting homunculus. The problem is that the interpretive behavior of homunculi seems to be the very kind of behavior that mental representations were posited to explain at the molar level. If so, we are led to identify psychologically internal structures of homunculi whose function is to interpret the mental representations that homunculi make use of in interpreting molar level mental representations. However this simply pushes the problem back one more level, with similar concerns leading to indefinitely "lower" levels of interpreting sub-homunculi.

One possible conclusion that can be drawn here is that we ought to give up talk of mental representations in accounting for any human behaviors. This, though, seems an especially Draconian sort of conclusion. After all, as Andy Clark and Josefa Roribio rightly note, "[C]ognitive science, it has often seemed, is agreed on at least this: that at the heart of a scientific understanding of cognition lies one crucial construct, the notion of internal representation."[19] Thus, giving up talk of mental representations in accounting for significant portions of human behavior ought to be a last alternative, not a conclusion immediately embraced. Accordingly, at this point advocates of mental representation explanations often adopt the strategy implicit in Fodor's example of shoe tying — a decompositional analysis into simpler and simpler subsystems until a point is reached where the interpretive activity is so simple that it has a known computational instantiation. On such an account, one discharges representations at some low level of analysis as "ontologically otiose", and so integrates the higher-level representational explanations into a naturalistic view of the world. The problem is that while, as Marvin Minsky suggests, this is a promising strategy in the case of something like intelli-

––––––––––

[18] Connectionists, while they too hold that mental representations are symbolic structures, deny that they are language-like objects of computation. While, for purposes of exposition, I use the analogy of mental representations as linguistic symbols, the points made hold, *mutatis mutandis*, for connectionist construals of mental representations.

[19] Clark and Toribio, p. 401.

gence,[20] it will not work for interpretation. In explaining intelligence, the idea is that we take a complex intelligent activity, like shoe tying, and break it down into the interaction of several simpler activities. The key is that at each level of decompositional analysis the interaction of less intelligent subsystems will explain the more complex, higher level intelligent activity. As William Lycan writes, we will explain "the successful activity of one homunculus, not by positing a second homunculus within it that successfully performs that action, but by positing a *team* of several smaller, individually less talented and more specialized homunculi—and detailing the ways in which the team members cooperate in order to produce their joint or corporate output."[21] Of course if the intelligent activity were equally complex at all levels, then, arguably, no real explanatory work would have been done. However, as long as the homuncular subsystems do not duplicate the intelligent activity they were posited to explain, then there is nothing wrong with the strategy. This works in the case of intelligence just because intelligence, as suggested by Fodor's example, comes in degrees. However, interpretation, unlike intelligence, does not come in degrees. The interpretive activity of homuncular subsystems may be more specialized than the higher-level system whose interpretive activity they were posited to explain, but they are no less interpretive than the higher-level system. If there was no computational instantiation of interpretive activity at the molar level, then there is no reason to think that it will be forthcoming at some homuncular level, the result being that representationalist explanations seem question begging.

──────────

[20] Minsky, p. 17, writes:

> This book [*The Society of Mind*] tries to explain how minds work. How can intelli-gence emerge from nonintelligence? To answer that, we'll show that you can build a mind from many little parts, each mindless by itself.

Also see Cummins, 1983: pp. 91-93. As Harré notes, p. 10, the idea of subpersonal psychology — as the strategy of transferring "the apparent psychological truths of commonsense psychologies into the scientific mode by the neat device of transforming personal functions into mental organs (or, in cybernetic terms, processing modules)" — is nothing new. Harré considers how the idea shows itself in both medieval morality plays and Freudian psychodynamics. It's also an idea that finds expression in Plato's tripartite conception of the soul.

[21] Lycan, 1996: p. 80. As Sterelny, p. 33, writes, "[W]e must not explain a cognitive capacity by tacitly positing an inner person with that very same capacity." Similarly, Sober, p. 420, writes that advocates of homuncular accounts believe that "[P]ostulating little men is permissible, as long as the little men do not have the same full-blown abilities as the people they inhabit." Also see Dennett, 1981: p. 124, and Haugeland, 2000: pp. 113-114.

Although the decompositional analysis of interpretive activity seems to fail, the failure is suggestive. If we are to keep representational explanations and concede their observer-relative character, then what gives rise to the problem is the assumption that the interpretive activity is internal to the person whose behavior is explained using mental representations. The problem, in other words, is an implicit acceptance of the mind as a sort of "Cartesian Theatre".[22] On such a view, representations appear on stage, are interpreted by the self/mind on whose theatre they appear, and then are used to achieve the self/mind's goals. It is this picture of the mind and the role of representations that must be resisted. Accordingly, the thesis that I will explore in the balance of this paper is that the problems posed by mental representations can be resolved by accepting their observer-relative character,[23] but denying that their "interpreter and user" must always be an internal sub-system of the human being to whom they are attributed.

In filling out the details associated with this claim, it is useful to begin by identifying two different questions:

> **(Q1)** When does the complex pattern of human behavior warrant attributing mental states and activities to a human being?
>
> **(Q2)** On the assumption that a human being's behavior warrants attributing mental states and activities to the human being, under what circumstances, if any, is it justified to account for such states in activities in terms of representations?

With respect to the first question, two remarks by Ludwig Wittgenstein in volume II of his *Remarks on the Philosophy of Psychology* are suggestive of the line I will advocate. The first remark is:

> "Human beings think, grasshoppers don't." This means something like: the concept of 'thinking' refers to human life, not to that of grasshoppers. (§ 23)

The second remark is:

> "Human beings sometimes think." How did I learn what "thinking" means? — It seems I can only have learned it by living with people. — To be sure, one could imagine seeing human life in a film, or being allowed merely to observe life without participating in it. Anyone who did this would then understand human life as we understand the life of fish or even of plants. We can't talk about the joy and sorrow, etc., of fish. (§ 29)

---

[22] See Dennett, 1991: pp. 107ff, and Sanders, pp. 4- 6.

[23] Dennett, 1981: p. 121.

What these remarks suggest is that the answer to the first question can only be found by looking at the roles and activities of the human being within a social practice. As Clark says, we attribute mental states and activities "by throwing a kind of interpretative net over a whole body of behavior. And the mesh of that net is gauged to our particular interest in making sense of behavior."[24] The interpretative net imposes a "holistic *interpretation* upon a large body of behavior in an environmental context"[25] with the intent of making *as much sense as possible* of that behavior relative to the *interests* of the community making the attributions of mental states and activities. This means that no one can unilaterally understand what it means for instances of behavior to warrant the attribution of mental states and activities except by reference to the authority of securable communal assent on the matter.

Relative to the first question then, the pattern of behavior exhibited by a human being warrants attributing mental states and activities to the creature only if the behavior is such that the community making attributions of mental states and activities accepts the behavior to be of such a sort that the attribution of mental states and activities is warranted.[26] Human beings whose behaviors do not, in enough cases, accord in the relevant ways with those of the community, will not be a creature to whom the community will attribute mental states and activities.[27] This does not mean that mental states and activities are in some sense "unreal" or "illusory". The situation is analogous to that of a mountain for a physicist. In her or his *scientific*, physical description of the world, no mention will be made of objects under the description of 'mountain'. This does not mean that mountains do not exist, it only means that no reference will be made to them under the description of 'mountain' in a scientific, physical description of the world.[28] However, in their lives apart from their practices as physicists, mountains *as* "mountains" will continue to play important recreational and geographic roles.[29] Similarly, it may be that

----

[24] Clark, p. 49. Also see Heil, 1983: pp. 163, 195.

[25] Clark, p. 153.

[26] Wittgenstein, 1969b: §§ 140, 281, 610. Also see Wittgenstein, 1970: §§ 534, 567, Wittgenstein, 1978: VI, §§ 20 - 21, 32, 34, Caraway, pp. 311 - 312, Dilman, pp. 162ff, and Lovibond, pp. 54ff. In contrast see Budd, pp. 318ff, Davies, p. 56 and McGinn, chapter 1.

[27] See Gier, p. 61, Kripke, chapter 3 and Lovibond, p. 51. In contrast see Budd, pp. 318ff.

[28] As Searle, 1981: p. 416, says, tables, chairs, mountains and beaches may be displaced by physics in the sense that the laws of physics "do not mention objects under the description 'chair', 'table', 'mountain' or beach'", but it does not follow from this that tables, chairs, mountains and beaches do not exist.

[29] Cf. Wilkes, 1986: pp. 174ff. Wilkes uses the example of fences rather than mountains.

for people seeking explanations for neurophysical happenings in people no attributions of mental states and activities will be made. At the same time though, in their lives apart from such practices, attributions of mental states and activities will continue to play important and essential roles. For instance, it is difficult to imagine how we could make sense of blaming and praising, sneering, hinting and insinuating, joking and punning, and in Kathleen Wilkes' words, "a wide variety of other Austintatious performances"[30] without reference to mental states and activities. Such a view is no more ontologically instrumentalist or irrealist concerning mental states and activities than physics is concerning the existence of mountains. At the same time though, it does mean that often there will be no naturalistic necessary and sufficient conditions that must be satisfied in order for attributions of mental states and activities to be warranted.

Given that within the context of many ordinary human practices it is proper to attribute mental states and activities, why explain them by making reference to mental representations? This is what the second question asks. The representationalist's answer to the second question is that if the most plausible explanations of the mental states and activities attributed to human beings that capture all the common sense psychological generalizations wanted are explanations that posit representations, then the posit will be made. Following a suggestion by Wilkes, the idea is that "[W]hat we cite in an explanation is in part a function of what puzzles our audience. People can be puzzled by all sorts of things, and so what we cite will depend as much on what we think the audience assumes about the problematic activity … as on what is supposed to hold of the agent."[31] For instance, think again of Clyde. We want to explain Clyde's behavior in going to the kitchen to get the beer. We know that Clyde does not invariably have beer in the refrigerator and that the obstacles to getting to the refrigerator vary from day to day. Thus, we are let to posit representations in order to account for Clyde's behavior. Does Clyde "use" the representations? In one sense, yes. We say that Clyde has a variety of representations and explain his behavior in terms of the ways that these representations contentfully interact with desires and other representations. However, we can say this without supposing that there is something internal to Clyde, behind the representations, making use of them in the way that a puppeteer uses a puppet. The mistake comes in supposing that we must say that Clyde is somehow different and apart from the representations. This is analogous to supposing that there is a difference between the thinker and the thinking. What we should say is that the thinker is the thinking, that Clyde is his representations. What makes them representations is not, necessarily, that they are representations for some internal psychological structure — a

_____

[30] Wilkes, 1984: p. 347.
[31] Wilkes, 1986: p. 177.

homunculus — but that they are representations for those of us accounting for Clyde's behavior. Although we "use" the representations in accounting for Clyde's behavior, our use is not the same as Clyde's use.

Of course, none of this precludes the possibility of there sometimes being psychologically internal systems for whom the representations are representations. Just as the attribution of personal level mental states and activities is made so as to, in Dennett's words, *make as much sense as possible of* the behavior of the human being relative to the interests of a community making such attributions[32], so too this is the motivation for positing representation using subsystems. Thus, the natural question is whether there any "good reasons" for supposing that the "interpreters of representations" will ever be (homuncular) subsystems internal to the person to whom mental states and activities are attributed? The answer to this question is, I believe, "yes". There is a range of interesting cases where it is natural and helpful to attribute such internal interpreters of representations.[33] For example, Amélie Rorty writes that "a self that is a loosely organized system composed of relatively autonomous subsystems seems hospitable to the possibility of self-deception. Self-deception is demystified and naturalized, and even to some extent explained, if the elf is a complexly divided entity for whom rational integration is a task and an ideal rather than a starting point."[34] This is a view shared by such writers as Herbert Fingarette[35] and Donald Davidson. Concerning Fingarette, Mike Martin writes that "Herbert Fingarette suggested that a self-deceiver—indeed, each of us—is a community of subselves which are organized clusters of desires-attitudes-emotions-beliefs-purposes, each of which can be expressed in semi-independence from other clusters."[36] Similarly, psychological phenomena such as amnesia, multiple personality disorders[37], and introspection, to name only three, have often led people to posit inner, relatively autonomous subsystems as constituting the self. The important point is that it is an empirical question whether we need to posit internal interpreters of mental representations.[38] If it turns out that withholding

––––––––––

[32] Dennett, 1989: p. 91.

[33] As Heil, 1994: p. 112, notes, "[T]he concept of a divided mind has an ancient and honorable history, going back at least to Plato's account of the tripartite soul in the *Republic*." Heil goes on to write, p. 118, that contemporary theorists "appeal to multiple selves not only in accounts of self-deception, weakness of will, and self-control but in discussions of conflicts among desires and between long- and short-term interests as well."

[34] Amelie Rorty, p. 213.

[35] See, for example, Fingarette, pp. 85ff.

[36] Martin, p. 27. Dennett is committed to the even stronger thesis that the self is a fictional "center of narrative grammar".

[37] See Braude, chapter 6.

mental representations.[38] If it turns out that withholding attributions of internal interpreters of representations will allow better sense to be made of a human being's behavior relative to the interests of the community making attributions of mental states and activities, or *will not contribute* in any meaningful way to making sense of the human being's mental states and activities relative to the interests of the community making attributions of mental states and activities, then such attributions will be withheld.

Here two questions naturally suggest themselves. First, if a decomposition of molar level mental states and activities occurs, must the decomposition continue because the explanation makes reference to the mental states and activities of homuncular subsystems? Second, if either there is no decomposition of the personal level mental states and activities, or there is decomposition but the decomposition stops at some "lower" level, isn't representational psychology question-begging?

We have already seen how an answer to the first question runs. Specifically, the answer depends on whether decomposing molar level mental states and activities into simpler, more specialized homunculi to whom mental states and activities are attributed, will contribute to making sense of the human being's mental states and activities. If the answer is 'yes', then molar level mental states and activities ought to be decomposed. At this next point, if continuing the decomposition of the mental states and activities will *further* contribute to making sense of the human being's mental states and activities, then the decomposition ought to continue. It is only when the stage is reached when nothing of further explanatory value is gained that the decomposition should stop. When this point is reached, the justifications for positing representation-using subsystems have been exhausted. The idea is not that further decomposition cannot be carried out, but that nothing of "interest" is gained in continuing the decomposition. What has been reached is, as Wittgenstein says, a "psychological, not a logical terminus"[39] where, rather than continuing the decompositional analysis, the representationalist may say that this is simply how the human being behaves. Just where the "cut off point" occurs will vary from situation to situation. For example, R.A. Sharpe writes that:

> When I use the words 'He is angry', I may do so because I see straight off that the man is angry. I do not infer this from his behavior. I am not hypothesizing an *inner state* to explain his behavior.[40]

---

[38] See Richard Rorty, pp. 162-163, and Schwartz, pp. 1056ff.
[39] Wittgenstein, 1970: § 231.
[40] Sharpe, p. 381.

Here the cut off point occurs at the molar level. On the other hand, cases like self-deception and multiple personality disorders suggest that sometimes the decompositional analysis may go quite deep. Having said all this, it is, of course, still an open and interesting question about how these mental states and activities are realized in the biology of the human being, and how they are affected by changes in the biology. This, though, is a different kind of question.

What then of the second question about whether representational psychology is question begging? In reality, there are two questions here. First, there is the question of whether any non-question begging explanation of mental states and activities must be one that explains the mental *in terms of* the non-mental. As Michael Tye notes, some conceptions of naturalism claim that answers to questions about mental states must be framed in a vocabulary that is "entirely non-mental (for example, behavioral or functional)."[41] Second, there is the question of whether stopping the regress of interpreters of representations with public interpreters begs the question of how these public interpreters are able to interpret representations. The problem is this: If a given subject's possessing representational mental states is contingent upon the existence of a community of interpreters of representations, then presumably the same is true of the representational mental states of the members of the community. Thus, there must be a second community of interpreters to anchor the representational mental states of the first community, and so on *ad infinitum*. It seems to follow that either there is a vicious infinite regress of interpreters of representations, or stopping the regress at the community level is question begging.

To understand how the representationalist can answer these two interrelated questions, it is necessary to distinguish three different kinds of explanations[42]:

(a) Explanations of high-level (complex) mental states and activities in terms of lower-level (less complex) mental states and activities;
(b) Explanations of mental states and activities in terms of non-mental states and activities;
(c) Explanations of high-level (complex) non-mental states and activities in terms of lower-level (less complex) non-mental states and activities.

It is true that if one recognizes *only* explanations of the form **(b)** or **(c)**, then explanations of form **(a)** are question begging. If the only way to genuinely explain mental states and activities is in terms of some non-representational vocabulary, then **(a)** form explanations are *not* genuinely explanatory. How-

---

[41] Tye, p. 422.
[42] Richard Rorty, p. 163, seems to conflate explanations of type (a) and (b).

ever, this is where to recall Wittgenstein's remark from the beginning of the paper—sometimes an explanation consists in "the understanding which consists in seeing connections…"[43] As implied above, the goal of representationalist psychology is not to account for how mental states and activities are instantiated as biological structures. Rather, the goal is to make as much sense as possible of certain kinds of psychological behaviors relative to the community interested in those behaviors. This runs counter to Paul Churchland for whom "… the primary purpose of folk-psychological talk is to fix on neurophysiologically sound states of the head so as to facilitate the prediction and explanation of others' … bodily movement"[44]. Representationalist psychology is not, as argued by Patricia Churchland and Paul Churchland, a stone-age relative of more respectable contemporary scientific theories. Instead, what we have are what Wittgenstein would call two different language games; one language game concerned with non-representational states and activities (for example, physics, biology, etc.), and another language game concerned with the attribution of mental states and activities, and making representational explanations on the basis of those attributions. Following out a remark of Wittgenstein's in *The Brown Book*, David Bloor puts it this way:

> [The language game associated with representationalist psychology] consists in making our actions intelligible by relating them to an accepted pattern of conduct and explanation. These patterns are known to all competent social actors and so they are readily available for the purposes of explanation and justification.[45]

Because the representationalist is offering explanations within the mental realm, explanations of form **(a)** are not question begging. To suppose otherwise is tantamount to saying that explanations of form **(c)** are question begging because they explain high-level non-representational states and activities in terms of low-level non-representational states and activities. But this is the wrong conclusion to draw. As long as the explanations are explanations of non-representational states and activities, then explanations of form **(c)** are perfectly appropriate, and *mutatis mutandis*, explanations of form **(a)** are also perfectly appropriate.

In addition, often the representationalist is not interested in providing explanations of high-level mental states and activities in terms of lower level mental states and activities. In many cases the attributions of mental states and activities are simply the results of the use of our practical skill of knowing how to deal with the world; "how to interact with the human and non-human

---

[43] Quoted in Monk, p. 451.
[44] Clark, p. 201n1.
[45] Bloor, p. 72

environment."[46] In such cases the attributions *show* a practical skill and do not present themselves as candidates for decompositional explanations. Indeed, many times such attributions present themselves as candidates for decompositional explanations *only* because our practical skill of knowing how to deal with the world either fails us or is thwarted. Thus, recognizing that the language games of making use of a mental vocabulary are different from those making use of a non-mental vocabulary leads to a rejection of the claim that explanations of form **(a)** are question begging.[47]

What now of the second problem—the problem of an infinite regress of communities of interpreters? Here the answer is that even if there is in some sense a regress, it is not a vicious regress. This is really the insight provided by the answer to the first problem. It is true that making the interpretation of representations external to the human said to have the mental representations leaves the interpretative activity of the external community members unexplained. However, why should this be a problem? We have explained the mental states and activities of the interpreted human, and this is what we wanted to do. Of course a different question can now be asked; viz., what about the mental states and activities of the community members engaged in the interpretation? This too, though, has an answer. For any given member of the community, that person's interpretive activity is explained by positing mental representations that are interpreted as representations by other members of the community. To be sure, there is a kind of regress, but just as the regress is not vicious in coherence theories of epistemic justification, so too it is not vicious here. If there is something problematic here, it is that we seem left with mental states that have not been explained in terms of a non-mentalistic vocabulary. As has already been suggested, this kind of objection rests upon a conflation of two different, though not incompatible, language games. The regress is vicious only if one insists that any non-vicious explanation of mental states and activities must finally "terminate" in explanations making use of a non-mental vocabulary. It is this insistence that must be resisted.

Finally then, it is important to note that the sort of explanations that the version of representationalism advocated in this paper makes use of are not

_____

[46] Stoutland, p. 8.

[47] See Wittgenstein, 1972: pp. 20ff where a similar point is made concerning the relation between aesthetic and physiological (or, more broadly, physical) explanations. It is worth noting that such a conclusion is consistent with there *also* being explanations of type (b). Indeed, as Dennett, 1989: p. 74, writes, "[I]ntentional system theory is almost literally a black box theory". Also see Dennett, 1989: pp. 257ff. Presumably this view clashes with that of Fodor who, according to Putnam, 1988: p. 7, desires "to make belief-desire psychology 'scientific' by simply identifying it outright with computational psychology."

causal explanations modeled on the purported causal explanations of physics.[48] Does this mean that representationalist psychological explanations are not causal explanations? The answer to this question is, I believe 'no'. However, without some elaboration this answer is misleading — misleading because the question is ambiguous. If the question is whether the 'because' in explanations such as "Jason bought the concert ticket because he wanted go to the Korn concert" is causal, then my answer is 'yes'. If the question is whether this kind of causal explanation is the same sort of causal explanation as found in the "paradigmatic" natural science of physics, then my answer is no. Yes, the explanation does support a host of counterfactuals, but no, it is not an explanation that instantiates a physical law. Rather the explanation reflects the fact that we often use "causal explanations" for macroscopic objects in ordinary life that do not lend themselves to either decompositional or eliminative analysis in terms of strict, nomological causal explanations. For instance, it is perfectly appropriate and useful to say in the context of a baseball game that the batter hitting the pitched ball caused the ball to fly out of the stadium. At the same time, it would be extremely surprising whether there is some physical law that is uniquely satisfied by this event. An analogous story can be told in the case of representations. If the question is whether at least some, perhaps properly regimented, representational explanations will turn out to be causal explanations in the more narrow physical sense of causal explanation, then my answer is that this is a question that cannot be answered *a priori*. Because representationalist explanations depend upon a specification of the interests of community making the explanations, then it is possible (though I think unlikely) that some representationalist psychological explanations will be more narrowly causal. Which representationalist explanations may turn out to be more narrowly causal? The answer to this question will depend upon an empirically discovered answer to the question: Are there psychological predicates based upon the interests of a particular linguistic community making attributions of mental states and activities that pick out natural kinds? If there are such predicates, then it is possible that at least some part of representationalist psychology could be absorbed into a science making use of a non-mental vocabulary. In contrast, if there are no such predicates,[49] then there will be no such absorption, and no representationalist psychological explanations will be narrowly causal.[50] In either case, representationalist psychology and non-representational science

---

[48] In contrast see Churchland, 1989: chapter 6.

[49] As suggested by Wittgenstein, 1970: §§ 608ff. Also see Wittgenstein, 1979: p. 180e. Whether there exist psychological predicates that pick out natural kinds is, I believe, presently unknown. In this connection see Graham and Horgan, pp. 73ff.

[50] See Davidson, pp. 207 - 227, 229 - 239, and 245 - 259.

are *not* in competition with each other.[51] It is the tendency of scientism to see only micro-level physical explanations as genuine that must be resisted. Once this is done, a significant portion of human behavior can be explained representationally "in terms that are compatible with the broader, physicalistic view of nature provided by the natural sciences."[52]

Andrew Ward
School of Public Policy, 685 Cherry Street
Georgia Institute of Technology
Atlanta, GA, 30332-0345, USA
andrew.ward@pubpolicy.gatech.edu

References

Bloor, David, *Wittgenstein: A Social Theory of Knowledge* (New York: Columbia University Press, 1983).

Braude, Stephen E., First Person Plural: Multiple Personality and the Philosophy of Mind (Lanham, MD: Rowman and Littlefield, 1995).

Budd, Malcolm, "Wittgenstein on Meaning, Interpretation and Rules," *Synthese*, v. 58, n. 3 (March, 1984), pp. 303-323.

Caraway, Carol, "Criteria and Circumstances, " *The Southern Journal of Philosophy*, v. 22, n. 3 (Fall, 1984), pp. 307-316.

Churchland, Paul, *A Neurocomputational Perspective* (Cambridge, MA: Bradford Books/MIT Press, 1989).

——, *Scientific Realism and the Plasticity of Mind* (Cambridge: Cambridge University Press, 1979).

Clark, Andy, *Microcognition* (Cambridge, MA: Bradford Books/MIT Press, 1991).

Clark, Andy and Toribio, Josefa, "Doing Without Representing?," *Synthese*, v. 101, n. 3 (December, 1994), pp. 401-431.

Crick, Francis, The Astonishing Hypothesis: The Scientific Search for the Soul (New York: Charles Scribner's Sons, 1994).

Cummins, Robert, *Meaning and Mental Representation* (Cambridge, MA: Bradford Books/MIT Press, 1989).

——, *The Nature of Psychological Explanation* (Cambridge, MA: Bradford Books/MIT Press, 1983).

——, "What Can Be Learned From *Brainstorms*," *Philosophical Topics*, v. 12, n. 1 (Spring, 1981), pp. 83-92.

---

[51] See Wittgenstein, 1979: § 402, Wittgenstein, 1969a: pp. 57 - 59, and Hallett, p. 625.
[52] Stich, p. 246.

Davidson, Donald, *Essays on Actions and Events* (Oxford: Clarendon Press, 1982).

Davies, Stephen, "Kripke, Crusoe and Wittgenstein," *Australasian Journal of Philosophy*, v. 66, n. 1 (March, 1988), pp. 52-66.

Dennett, Daniel, *Brainstorms* (Cambridge, MA: Bradford Books/MIT Press, 1981).

——, "Cognitive Ethology: Hunting for bargains or a Wild Goose Chase," in *Brainchildren: Essays on Designing Minds* (Cambridge, MA: Bradford Books, 1998), pp. 307-322.

——, *Consciousness Explained* (Boston: Little, Brown and Company, 1991).

——, *The Intentional Stance* (Cambridge, MA: Bradford Books/MIT Press, 1989).

——, "The Interpretation of Texts, People and Other Artifacts," *Philosophy and Phenomenological Research*, v. 50, supple. (Fall, 1990), pp. 177-194.

——, "Why Believe in Belief?," *Contemporary Psychology*, v. 30, n. 12 (November, 1985), p. 949.

Dilman, Ilham, "Wittgenstein on the Soul," in *Understanding Wittgenstein*, edited by Godfrey Vesey (Ithaca, NY: Cornell University Press, 1976), pp. 162-192.

Dretske, Fred, *Explaining Behavior: Reasons in a World of Causes* (Cambridge, MA: Bradford Books/MIT Press, 1988).

Fingarette, Herbert, *Self-Deception* (Berkeley: University of California Press, 2000).

Fodor, Jerry, *The Language of Thought* (Cambridge, MA: Harvard University Press, 1979).

——, *RePresentations* (Cambridge, MA: Bradford Books/MIT Press, 1981).

——, *Psychosemantics* (Cambridge, MA: Bradford Books/MIT Press, 1987).

Fodor, Jerry, and LePore, Ernest, *Holism: A Shopper's Guide* (Oxford: Basil Blackwell, 1992).

Gier, Nicholas, "Wittgenstein, Intentionality and Behaviorism, " *Metaphilosophy*, v. 13, n. 1 (January, 1982), pp. 46-64.

Graham, George, and Horgan, Terence, "How to be Realistic about Folk Psychology," *Philosophical Psychology*, v. 1, n. 1 (1988), pp. 69-81.

Hallett, Garth, *A Companion to Wittgenstein's "Philosophical Investigations"* (Ithaca, NY: Cornell University Press, 1977).

Harré, Rom, *Personal Being* (Cambridge, MA: Harvard University Press, 1984).

Haugeland, John, *Artificial Intelligence: The Very Idea* (Cambridge, MA: Bradford Books, MIT Press, 2000).

——, "Mind Embodied and Embedded," in *Having Thought: Essays in the Metaphysics of Mind* (Cambridge, MA: Harvard University Press, 1998), pp. 207-237.

——, "Representational Genera," in *Having Thought: Essays in the Metaphysics of Mind* (Cambridge, MA: Harvard University Press, 1998), pp. 171-206.

——, "The Intentionality All-Stars," in *Having Thought: Essays in the Metaphysics of Mind* (Cambridge, MA: Harvard University Press, 1998), pp. 129-170.

——, "The Mother of Intention," *Nous*, v. 16, n. 4 (November, 1982), pp. 613-619.

Heil, John, "Does Cognitive Psychology Rest on a Mistake?," *Mind*, v. 90, n. 359 (July, 1981), pp. 321-342.

——, "Going to Pieces," in *Philosophical Psychopathology*, edited by George Graham and G. Lynn Stephens (Cambridge, MA: Bradford Books/MIT Press, 1994), pp. 111-133.

——, *Perception and Cognition* (Berkeley: University of California Press, 1983).

Horgan, Terence, "Naturalism and Intentionality," *Philosophical Studies*, v. 76, nos. 2-3 (December, 1994), pp. 301-326.

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

Lovibond, Sabina, *Realism and Imagination in Ethics* (Minneapolis, MN: University of Minnesota Press, 1983).

Lycan, William, "The Continuity of Levels of Nature," in *Mind and Cognition: A Reader*, edited by William Lycan (Oxford: Basil Blackwell, 1996), pp. 77-96.

——, "Dennett's Instrumentalism," *Behavioral and Brain Sciences*, v. 11, n. 3 (September, 1988), pp. 518-519.

——, *Judgement and Justification* (Cambridge: Cambridge University Press, 1988).

McGinn, Colin, *Wittgenstein on Meaning* (Oxford: Basil Blackwell, 1984).

Martin, Mike W., *Self-Deception and Morality* (Lawrence, KS: University of Kansas Press, 1986).

Minsky, Marvin, *The Society of Mind* (NY: Touchstone Books, 1986).

Monk, R., Ludwig Wittgenstein: The Duty of Genius (NY: Free Press, 1990).

Putnam, Hilary, *Reason Truth and History* (Cambridge: Cambridge University Press, 1981).

——, *Representation and Reality* (Cambridge, MA: Bradford Books/MIT Press, 1988).

Rorty, Amelie, "The Deceptive Self: Liars, Layers, and Lairs," in *Mind in Action: Essays in the Philosophy of Mind* (Boston: beacon Press, 1988), pp. 212-228.

Rorty, Richard, "Wittgenstein, Philosophy and Empirical Science," *Philosophical Studies*, v. 31, n. 3 (March, 1977), pp. 151-172.

Sanders, John T., "An Ecological Approach to Cognitive Science," *The Electronic Journal of Analytic Philosophy*, v. 4 (Spring, 1996), pp. 1-12.

Schwartz, Robert, "'The' Problem of Representation," *Social Research*, v. 51, n. 4 (Winter, 1984), pp. 1047-1064.

Searle, John, "Analytic Philosophy and Mental Phenomena," in *Midwest Studies in Philosophy, Volume 6: The Foundations of Analytic Philosophy*, edited by Peter French, Theodore Uehling, Jr., and Howard Wettstein (Minneapolis, MN: University of Minnesota Press, 1981), pp. 405-423.

——, "Is the Brain a Digital Computer?," *Proceedings and Addresses of the American Philosophical Association*, v. 64, n. 3 (November, 1990), pp. 21-37.

——, "The Realist Stance," *Behavioral and Brain Sciences*, v. 11, n. 3 (September, 1988), pp. 527-529.

Sharpe, R.A., "The Very Idea of a Folk Psychology," *Inquiry*, v. 30 (1987), pp. 381-393.

Sober, Elliott, "Why Must Homunculi Be So Stupid?," *Mind*, v. 91, n. 363 (1982), pp. 420-422.

Stich, Stephen, "What is a Theory of Mental Representation?," *Mind*, v. 101, n. 402 (April, 1992), pp. 243-261.

Sterelny, Kim, The Representational Theory of Mind: An Introduction (Oxford: Basil Blackwell, 1990).

Stoutland, Fred, "Against Folk Psychology," (1992), unpublished.

Tye, Michael, "Naturalism and the Mental," *Mind*, v. 101, n. 403 (July, 1992), pp. 421-441.

Wilkes, K.V., "Nemo Psychologus nisi Physiologus," *Inquiry*, v. 29, n. 2 (June, 1986), pp. 169-185.

——, "Pragmatics in Science and Theory in Common Sense," *Inquiry*, v. 27, n. 4 (December, 1984), pp. 339-361.

——, "The Relationship Between Scientific Psychology and Common-Sense Psychology," *Synthese*, v. 89 (1991), pp. 15-39.

Wittgenstein, Ludwig, *The Blue and Brown Books* (Oxford: Basil Blackwell, 1969a).

——, *Lectures and Conversations on Aesthetics, Psychology and Religious Belief*, edited by Cyril Barrett (Berkeley: University of California Press, 1972).

——, *On Certainty*, edited by G.E.M. Anscombe and G.H. von Wright, translated by Denis Paul and G.E.M. Anscombe (New York: Harper and Row Publishers, 1969b).

——, *Philosophical Investigations*, third edition, translated by G.E.M. Anscombe (New York: Macmillan Publishing Company, 1979).

——, *Remarks on the Foundations of Mathematics*, revised edition, translated by G.E.M. Anscombe (Cambridge, MA: MIT Press, 1978).

——, *Remarks on the Philosophy of Psychology, v. 1*, edited by G.E.M. Anscombe and G.H. von Wright, translated by G.E.M. Anscombe (Chicago: The University of Chicago Press, 1980a).

——, *Remarks on the Philosophy of Psychology, v. 2*, edited by G.H. von Wright and Heikki Nyman, translated by C.G. Luckhardt and M.A.E. Aue (Chicago: The University of Chicago Press, 1980b).

——, *Zettel*, edited by G.E.M. Anscombe and G.H. von Wright, translated by G.E.M. Anscombe (Berkeley: University of California Press, 1970).