Connectionism and the Mind-Body Problem: Exposing the Distinction Between Mind and Cognition.

Tim van Gelder

appeared in *AI Review;* also in L. Niklasson & M. Boden (Eds.), *Selected Readings of the Swedish Conference on Connectionism 1992.* Ellis Horwood.

Abstract. Does PDP-style connectionism imply that there are no such things as beliefs? Ramsey, Stich and Garon (1991) have argued that it does, but their argument depends on a particular construal of the ontology of beliefs, according to which beliefs are discrete inner causal entities. In this paper I situate their argument within the broader issues of the clash of the scientific and manifest images of the world, and the ontological mind-body problem. Their way of understanding belief places them in a long tradition of philosophers who conceive of mind and cognition as the same thing, a tradition that includes people as diverse as Descartes, Fodor, Churchland, and virtually all current cognitive scientists and cognitive neuroscientists. An alternative perspective on the nature of belief was championed by Ryle and is currently exemplified by Dennett's neo-instrumentalism. I argue that PDP-style connectionism, rather than implying that there are no beliefs, implies that the construal of the ontology of belief within the mind-as-cognition framework is mistaken. Beliefs are not discrete inner causal entities. Mind is not ontologically homogeneous; rather, like an economy, it is made up of a diversity of entities of very different ontological kinds interrelated in complex ways. Hence cognitive science is not the science of the mind; it is the science of cognition, which is only one aspect of mind.

0. Introduction

As a research paradigm within cognitive science, Connectionism is surely exciting, fertile and provocative. It has expanded the range of methods and concepts available for the study of cognition. It is producing new kinds of explanations of psychological phenomena. It has forced reconsideration of a wide range of foundational issues in the philosophy of cognitive science. Some have even gone so far as to claim that it has dramatic implications for one of the oldest issues in philosophy, the mind-body problem.

At first glance, this should not be surprising. After all, cognitive science is, in Howard Gardner's succinct encapsulation of its official self-image, *The Mind's New Science* (Gardner 1985). Connectionism, as a branch of cognitive science, is consequently an exciting new way of studying the *mind*, and so must surely have consequences for the relationship between the mind and the body. And, in fact, a number of philosophers have claimed that Connectionism supports *Eliminativism*, the doctrine that some or all of the entities of mind as ordinarily understood simply don't exist. In this paper I will discuss Ramsey, Stich & Garon's (1991) careful elaboration of the thesis that, if a certain kind of Connectionism is right, then we don't have beliefs. In the background, however, is Paul Churchland's more general and dramatic version of Eliminativism, according to which virtually all of mind as traditionally conceived is either being shown not to exist, or is being radically reconceived, by the advance of Connectionism and cognitive neuroscience (Churchland 1988).

I will argue that Connectionism only implies Eliminativism within the context of a certain broad picture of mind, which can be called the "Mind as Cognition" doctrine. Connectionism can therefor be taken as implying, not Eliminativism, but rather that the Mind as Cognition doctrine itself is mistaken, a position which is philosophically attractive for a variety of other reasons. Taking this route, Connectionism does indeed have some dramatic implications for the mind-body problem, though ones very different from those drawn by Ramsey, Stich and Garon.

1. Manifest and Scientific Images of Reality and the Mind-Body Problem.

In discussing these issues it is useful to begin by taking over Wilfred Sellar's distinction between two quite different overall pictures of man and the world, the manifest and scientific images (Sellars 1962). The manifest

image is embedded in all our ordinary ways of dealing with the world around us, each other, and ourselves. Within the manifest image, the world made up of such familiar things as trees, lightning, tables, heat, colors, weight, people, beliefs, moods, purposes, skills, conversations, promises, humor, laws, and so forth. The scientific image is the world as it is portrayed by the "hard" sciences - physics, mechanics, chemistry, biology, and so forth. Both images purport to truly depict the world, and to some extent they are in conflict. Thus, in the manifest image there is a brown table in front of me as I write this. However, brownness and tables figure nowhere in hard science, which would provide a variety of very different descriptions of the object in front of me, depending on whether you consulted physics, chemistry, or biology.

One standard response to the clash of images has been to take the scientific image as providing the true picture, the only description of the way the world really is. The manifest image is regarded as a clumsy, naive picture of the world, either a non-scientific picture, or perhaps an early but inadequate attempt at a scientific picture. What then becomes of all the entities belonging to the manifest image? There are at least four basic strategies for accounting for the status of manifest entities in relation to the scientific image:

The first, *incorporation*, is to loosely indicate some kind of identity between the manifest entity and some entity or construction of entities from the scientific image, as when we casually accept that lightning is a movement of electrons.

The second, *elimination*, is to accept that the scientific image shows that the manifest entity does not, in fact, exist at all. For example, Paul Churchland has argued that contemporary science has shown that there is in fact no such thing as *temperature* as we ordinarily understand it. This is because science has uncovered three rather different concepts in the vicinity of the everyday concept of temperature, none of which is similar enough to the everyday concept to be identified with it.

The third strategy, *interiorization*, is to pull the manifest entity back into the mind. It is designated not an actual feature of the world in its own right, but rather just the way that world appears in the mind of an observer. Thus the brownness of the table is not itself "out there", but rather is only the way that certain physical phenomena that science describes in quite different terms (e.g., photons and wavelengths) appear to us.

The fourth response, *exclusion*, works by excluding the manifest entity from the domain of science altogether. The manifest entity certainly exists, and not merely "in the mind", but cannot plausibly be identified with anything yet describable in the scientific image. An example of this would be a law, or body of law, such as a country's constitution; it is something about which science, in anything like its current form, has nothing to say.

One consequence of the development of a scientific image of the world, and the consequent friction between the manifest and scientific images, has been exacerbation of the traditional ontological mind-body problem. That is the problem of determining what kind of things mind and mental entities are, and describing how they fit into the rest of the natural world. This problem was made more difficult by the emergence of the scientific image in two ways. First, it became the attempt to reconcile mind as conceived in the manifest image with the new scientific descriptions of the physical world. Any solution had to somehow bridge the gap, not just between mind and world, but between the manifest and scientific images. Second, that emerging scientific image was one which had virtually nothing useful to say about mind. The mind-body problem became that of accounting for the place of mind in a world that seemed to have no room for it.

The scientific image is in constant flux. In recent decades it has been developing especially rapidly as a whole new branch of science has come into play - i.e., cognitive science, here understood very broadly so as to include cognitive neuroscience. If we accept the standard description of cognitive science as the science of the *mind*, then we are witnessing the development of a scientific image of mind. In response, the traditional mind-body problem has mutated into two problems, both of which are presumably easier to solve: first, locating the manifest image of mind with respect to the newly emerging scientific description of mind, and second, locating

the scientific description of mind correctly within the scientific worldview as a whole.

Philosophers of mind and of cognitive science have been engaged in both projects, often simultaneously. In what follows, I will be focusing on the former, and in particular on the question: what is going to become of all the familiar mental entities from the manifest image with the development of the scientific image - or partial image - of mind? In particular, what will happen to these entities with the development of a *connectionist* image of mind? What will happen to all the familiar components and manifestations of mentality, such as beliefs, sensations, images, emotions, moods, intelligence, wit, sensitivity, empathy, humor, and so forth? Will they be incorporated, eliminated, interiorized, or excluded?

Clearly, interiorization - relegating back into the mind - is not an option at this stage. It might be possible to explain away some feature of the manifest image of the *world*, such as brownness, by pulling it back into the mind, but one cannot explain away some feature of the *mind* that way.

Much of mainstream cognitive science is based on a strategy of incorporation with respect to at least some of the entities familiar from the manifest image. As a matter of practice, it takes for granted the existence of manifest entities such as beliefs, desires, sensations, and images, and constructs models of cognitive functioning which incorporate states, structures and processes corresponding to them. Some cognitive scientists go so far as to insist that incorporation is, effectively, mandatory. In particular, they think that the ontological *bona fides* of entities such as beliefs and desires is established by their utility within the manifest image, and, they argue, any cognitive science that is going to stand a chance of explaining that utility must incorporate entities closely corresponding to beliefs and desires in its mechanistic models (see e.g., Fodor 1987).

Much connectionist research, on the other hand, raises the possibility of *elimination* with respect to those very same entities. Connectionism is developing a scientific picture of mind in which (by and large) the mental entities of manifest image simply do not figure; nor is there anything in connectionist models with which such entities might be even loosely identified. This suggests that those entities simply do not exist. This is the broader context within which Ramsey, Stich and Garon make their case. What they argued, in effect, was that in a certain kind of connectionist work, the gap between the emerging scientific image of mind and the manifest image is so great that certain manifest entities - namely, beliefs - are being shown not to exist.

There is, of course, one more possibility: that of *exclusion*. It may be that some aspects or manifestations of mind, as it figures in the manifest image, are just so different in ontological status from the kinds of things that cognitive science in general and connectionism in particular are studying, that those sciences are simply not operating in the same domain, and so do not stand either to incorporate or to eliminate those aspects. For example, it is often claimed that science, as we know it, has nothing to say that is even remotely relevant to the problem of *qualia*, the qualitative characters of our conscious experiences, and so the scientific image in its current form neither includes nor precludes qualia. In what follows, I will suggest that, whether or not this is true for qualia, it probably *is* true in the case of *belief*. When we get an accurate bearing on the kind of things that beliefs actually are, as they figure in the manifest image, we can see that whether the scientific image contains discrete entities inside the head corresponding to beliefs is irrelevant to the issue of whether we have beliefs. The scientific image of mind can neither incorporate nor eliminate beliefs; rather, they are outside the domain of the scientific image as it is currently developing.

The next section will also do some philosophical stage-setting, though from a different perspective. It will briefly trace some of the history of the mind-body problem, offering a highly selective sketch some of the key issues and theories leading up to the point where Ramsey, Stich & Garon advanced their arguments concerning the eliminativist implications of connectionism.

2. The Mind-Body Problem: A Brief History

As pointed out already, much of the difficulty of the mind-body problem was due to the absence of anything

like mind from the scientific image. Consequently, discussions of the problem have tended to focus on the question: is the mental just some part or aspect of the physical world, or is it something fundamentally distinct? This is the clash between materialism and dualism. The seventeenth century English philosopher Thomas Hobbes is a useful representative of the former view; his contemporary René Descartes is of course the most famous representative of dualism.

Hobbes began his classic treatise on political philosophy, *Leviathan*, by attempting to show that humans are nothing but complex organizations of matter, and that all mental phenomena are some kind of internal physical motion. His first move was to explain *sensation*, as follows:

The cause of sense, is the external body, or object, which presseth the organ proper to each sense...which pressure, by the mediation of the nerves, and other strings and membranes of the body, continued inwards to the brain and heart, causeth there a resistance, or counter-pressure, or endeavour of the heart to deliver itself, which endeavour, because *outward*, seemeth to be some matter without. And this *seeming*, or *fancy*, is that which men call *sense*... Neither in us that are pressed, are they any thing else, but divers motions; for motion produceth nothing but motion... (Hobbes 1962 p.21).

Then, in a series of rather implausible reductive maneuvers, he tried to account for all other mental phenomena as some kind of construction from sensations. Thus both *imagination* and *memory* are forms of "decaying sense", and *understanding* a form of imagination resulting from the perception of signs; *passions* are sensations which are "interior beginnings of voluntary motions", i.e., which cause actions. In principle, all human behavior can be explained in terms of the causal mediation of these internal physical motions between physical stimulation of the sense organs and motions of the body.

Descartes was one of the leading anatomists and physiologists of his time, and, like Hobbes, aimed to explain as much of animal and human behavior as possible in mechanistic terms. He believed that animal and human bodies were sophisticated machines made of bone, muscles and nerves, and devoted considerable effort to demonstrating how many different kinds of behavior arise solely from the structure and functioning of these machines. He believed all animal behavior could be explained in these terms; animals were nothing but machines. Not all *human* behavior could be explained this way, however. In particular, he argued that no machine, no matter how complex, could either *talk* as people do or exhibit such a wide *range* of skillful behaviors. These behaviors need the cooperation of something else, something non-material; the obvious candidate was a *mind*. In a famous passage he characterizes the mind as "a thing which thinks...a thing which doubts, understands, affirms, denies, wills, refuses, which also imagines and feels." (1911 p.153). Descartes thus believed that the universe contained at least two fundamentally different kinds of entities - material things, whose essence was to be spatially extended, and mental things, whose essence was to think. A person is a conjunction of a body and a mind, which, despite their deep ontological difference, causally interact in perception and action. In perception, physical contact moves a nerve ending, and that movement travels all the way along the nerve to the brain, and in particular to the pineal gland, there generating a sensation in the mind. This sensation is a particular conscious event inserted into the stream of mental occurrences. Action is symmetrical with perception; first there is a conscious mental act, a volition, which causes a physical motion in the pineal gland, which is transmitted down the nerves to the muscles and causes them to contract appropriately.

In standard discussions of positions on the mind-body relationship, Hobbesian materialism and Cartesian Interactionist Dualism are regarded as occupying the extremes in a spectrum of options. While there are certainly fundamental differences between the two views, it is also important to see how much they had in common. Standing back from the details, we can see that they were both committed to the following deep structural features:

(1) *Mind as Manifest.* Hobbes' and Descartes' catalogues of mental phenomena are basically the same: mind is made up of occurrent episodes of perceiving, understanding, willing, dreaming, remembering, imagining, and so

forth. These catalogues are drawn directly from the manifest image. The problem for both philosophers was to explain the relationship to the physical world, and in particular to the body and brain, of the entities that figure in our everyday, non-scientific ways of describing ourselves and others in mentalistic terms.

(2) *Mind as Ontologically Homogenous.* Hobbes and Descartes each have their own relatively simple and unified *ontology* of mind. Though there is a diversity of particular mental phenomena, they are all of the same basic ontological kind: for Descartes, modifications of mental substance, and for Hobbes, physical motions within the brain.

(3) *Mind as Internal.* For both, mind is something "internal". For Hobbes, the mental is internal to the skull, and as a matter of practical fact unobservable by anyone else. For Descartes, mind is internal in an even stronger sense: the happenings within a mind are *in principle* observable only by that mind itself. As Gilbert Ryle was later to put it, minds are "insulated fields".

(4) *Mind as Causal Underpinning of Behavior.* Both Hobbes and Descartes believed that mind is an essential part of the full causal explanation of human behavior.

In short, both Hobbes and Descartes believed that *all familiar mental entities are internal causal underpinnings of behavior*. This is the essence of a broad perspective that I call the "Mind as Cognition" doctrine. Their positions can be regarded as two different ways of elaborating the basic structure of the Mind as Cognition doctrine, one which attempts a purely materialist story and one which finds the need for non-material intermediaries.

It was not until the twentieth century that the basic assumptions of the Mind as Cognition picture were identified and challenged. In his 1949 book *The Concept of Mind* Gilbert Ryle attacked the Cartesian position in particular, labeling it "the myth of the ghost in the machine". But while Cartesian dualism was his primary target, most of his criticisms didn't depend upon his target being a form of dualism, and so were equally effective against any position that fell within the broad scope of the Mind as Cognition framework. Unfortunately, the nature of Ryle's quasi-behaviorist alternative was difficult to articulate, and it came to be caricatured as the view that all mental entities are a matter of external behavior. Such a position was thought to be obviously absurd and so Ryle's views were neglected.

In the 1950s, Hobbesian materialism was resurrected in the form of the Identity Theory, the view that every mental entity is some aspect of the brain and its functioning (e.g., Place 1956). This doctrine has an initial ring of plausibility, but the subsequent history of the mind-body problem has amounted to a series of retreats from any strong version.

Initial formulations of the Identity Theory are standardly taken to have postulated *type*-identities: every *kind* of mental state corresponds to a distinctive *kind* of neurochemical state.[1] Thus, there is some one kind of neural state for being depressed, one for feeling a sharp pain in the toe, and one for believing that there are grizzly bears in Yellowstone National Park. For various reasons, however, the type identity thesis seemed too strong. The most well-known is the "multiple-realization" argument: surely computers or Martians can also believe there are grizzlies in Yellowstone, even if their "brains" are physically constituted in some very different way than ours. For this reason we should not expect that there is any one type of brain state for each kind of mental state. Put differently, it was recognized that the mental and the physical amount to two very different ways of grouping things into *kinds*. Materialism could be saved by retreating to a "token-identity" thesis: every individual "token" of a mental kind is some physical entity; in us, it is a feature of the brain. Different tokens of the same mental kind might have very different physical realizations in different people or across species.

If the mental groups entities into kinds in some way differently than the physical, what principles underlie that grouping? This is really just the question: what makes a particular feature of the brain a sharp pain, or a belief that there are bears in Yellowstone? The standard answer is: its causal role. A state of my brain is a sharp pain

if it is caused by things like jabs and in turn causes me to say, do, and believe characteristic things. This is the core of *Functionalism*: all particular mental entities are physical states, and their mentality is a matter of their distinctive causal role.

The currently dominant form of this view is *Computational Functionalism*, which cashes out the notion of causal role in terms of an inner computational architecture and its processes. These processes might be computational in a quite strict sense - that is, the algorithmic manipulation of structured representations. Thus, Fodor for one holds that a belief that there are bears in Yellowstone is a brain entity which happens to be a structured representation of that proposition playing the right causal role in a classical computational cognitive architecture (Fodor 1987). Computational Functionalism need not be based on a classical conception of cognition, however; mental entities can be physical entities with causal roles characterized in terms of the kinds of computational processes found in connectionist networks, for example.

Computational Functionalism, broadly construed, is the closest we have to a contemporary orthodoxy on the mind-body problem. It is officially subscribed to by a significant portion of philosophers, and is held, mostly implicitly, by most cognitive scientists. It has however been challenged in a number of ways. Currently, the most prominent challenge comes from *Eliminativism*, the doctrine that some or all categories of familiar mental entities simply do not exist (Churchland 1988).

How can such a radical position be defended? Typically, the first move is to argue that our ordinary ways of describing and explaining ourselves and others in mentalistic terms actually amount to the deployment of a primitive scientific theory, dubbed "folk psychology" because it both applies to and is utilized by everyday folk. The second move is to point out that, like any scientific theory, folk psychology is liable to be supplanted by more advanced theories in the same domain. It may happen that folk psychology is actually supplanted by a cognitive science which describes and explains human behavior in terms of states and processes which bear little or no resemblance to the beliefs, desires, sensations, emotions, etc. of folk psychology. If that happens, those entities will have been shown not to exist at all, in just the same sense that we now know that there is no such thing as phlogiston.

Computational Functionalism and Eliminativism are the contemporary representatives of two of the basic strategies for accounting for the status of manifest entities in light of the scientific image. Thus, Computational Functionalism is basically incorporationist: it believes that a fully developed science of human behavior is going to actually find the mental entities of the manifest image inside the head and characterize them in physical terms. Eliminativism, of course, opts for elimination.

Both Computational Functionalism and Eliminativism are, like Hobbesian materialism and Cartesian dualism, just alternative ways of working within the Mind as Cognition framework. Both positions take both the manifest image of mind and the scientific image of mind to be descriptions of the internal states and operations which are causally responsible for our behavior. The difference between them is simply over whether science will end up incorporating the familiar entities of the manifest image. The Computational Functionalist is confident that those entities will figure in the scientific story; the Eliminativist doubts their value and expects a mature Cognitive Science to dispense with them. The Computational Functionalist believes in the consistency of the four basic principles of the Mind as Cognition picture (Mind as Manifest, Mind as Ontologically Homogeneous, Mind as Internal, and Mind as Causal Underpinning of Behavior). The Eliminativist believes that they are inconsistent, and is prepared to reject the first (Mind as Manifest) as a way of preserving a commitment to the other three.

3. Connectionism and Eliminativism

This, in broad outline, is the philosophical stage onto which Connectionism enters. Connectionism is potentially significant for the mind-body problem because, as part of Cognitive Science, it is a way of examining the internal causal underpinnings of human behavior, and hence, a source of evidence in choosing between

Computational Functionalism and Eliminativism. As it happens, precious little of mind as we ordinarily know it figures in standard connectionist models. Connectionism therefore appears to throw its empirical weight behind Eliminativism.

This implication of Connectionism has been thoroughly worked out by Ramsey, Stich and Garon. They claim that Connectionism is "genuinely revolutionary" because it supports a "thoroughgoing Eliminativism about some of the central posits of common sense (or "folk") psychology" (p.199). In particular, they claim that Connectionism, if correct, shows that there are no such things as *beliefs*. This is because, if we look inside Connectionist models, we don't find anything that can be plausibly identified as individual beliefs. The relevant argument can be succinctly summarized as follows:

(1) Beliefs, as we ordinarily understand them, are functionally discrete, semantically interpretable internal states that play a causal role in the production of behavior.

(2) A certain important sub-class of Connectionist networks - namely, distributed, subsymbolic cognitive models - do not contain any such states.

(3) Connectionist networks of this kind properly describe the human cognitive architecture.

(C) Therefore, beliefs, as we ordinarily understand them, do not exist.

In fact, Ramsey, Stich & Garon don't present the argument quite this way because they don't wish to take a stand on whether the relevant kind of Connectionism is in fact correct. They advance the first two premises, and state the conclusion as a conditional: *if* Connectionism is correct, there are no beliefs.

Ramsey, Stich & Garon anticipate at least two kinds of responses to their argument. One is to insist that discrete entities corresponding to beliefs can in fact be found in the relevant kind of Connectionist networks - in other words, deny Premise 2. They consider various strategies for identifying such entities, and argue persuasively that they all fail: the entities thereby uncovered cannot in fact plausibly be called beliefs, as beliefs are described in Premise 1.

The second anticipated response amounts to the denial of Premise 3. They imagine someone responding that these kinds of Connectionist networks can't possibly be correct *precisely because* they don't incorporate entities corresponding to beliefs. This kind of response need not be simply begging the question, but to avoid doing that it must be backed up by some independent reason for supposing that any adequate cognitive model must incorporate entities corresponding to beliefs. Since Ramsey, Stich & Garon are only arguing for the conditional conclusion, they do not need to counter this kind of response.

4. Eliminativism and The Nature of Belief

There is, of course, another possible response to their argument: deny Premise 1. (Ramsey, Stich and Garon do not consider this strategy as a possible *response* to their argument, though they do spend quite a bit of space explaining and defending Premise 1.) This premise is an ontological claim; it tells us what kind of things beliefs, as we ordinarily understand them, actually are. If Premise 1 were false - if beliefs were not in fact discrete inner causal entities, but something quite different - then nothing would follow from the fact that there are no such entities in Connectionist networks. Thus, we can escape the eliminativist conclusion of their argument by providing an alternative account of the nature of beliefs.

Suppose, for example, that we accepted a view something like of Daniel Dennett (Dennett 1987). On that approach, a belief is simply an abstract state attributed to a whole system from a specialized perspective. Suppose you have a complex system and you want to be able to explain and predict its behavior. There are various ways you might go about this, but one is to take up what Dennett calls the "Intentional Stance". This means attributing to the system a reasonable set of beliefs and desires and the ability to act rationally given

those beliefs and desires. You can then predict what the system will do. It turns out that certain systems, such as people, behave in such a way that this predictive strategy works remarkably well, and in fact might the *only* effective strategy available. If you have reason to think that I like a pint of Watneys, and that I believe cold pints of Watneys can be obtained at the Irish Lion, then on this basis you can predict quite reliably where I'll be, and it is very unlikely that you have any other practical way to predict this.

The important thing for current purposes is the account of the ontology of belief. In this account, a belief is simply a state attributed to a system as a whole in the process of using the intentional strategy - a kind of token in a calculus of prediction. As Dennett puts it:

all there is to being a true believer is being a system whose behavior is reliably predictable via the intentional strategy, and hence *all there is* to really and truly believing that *p*...is being an intentional system for which *p* occurs as a belief in the best (most predictive) interpretation. (p.29)

Dennett has gone on to liken beliefs, ontologically, to entities known in the philosophy of science as "abstracta" - "calculation-bound entities or logical constructs", rather like the equator or centers of gravity. Crucially, beliefs are not discrete causal mechanisms inside the skull. Of course, there must be *some* complex mechanisms inside the head which are causally responsible for the regularities and subtleties of our behavior. Dennett usually doubts that anything about those mechanisms will correspond neatly to the beliefs and desires attributed in the intentional stance. Further, even if cognitive science and neuroscience were eventually to reveal internal items which did correspond to particular beliefs, what those sciences would have found are not the beliefs themselves but merely their causal underpinning: mechanisms which, in the context of the rest of the system's causal structure, are responsible for the patterns of behavior such that the system can be said to have beliefs.

If something like Dennett's account of belief is correct, then Premise 1 is false and Ramsey, Stich & Garon's eliminativist conclusion need not be accepted. Is it in fact correct? Are beliefs abstract states attributed from a specialized external perspective, or are they discrete internal causal mechanisms - or something else again? In my opinion, Dennett's account is much closer to the truth, though I can't right now offer a fully adequate ontology of belief. In the remainder of this section I will just lay out some general considerations which tend to support an externalist account.

(1) The problem here is to get clear on the ontological status of a kind of entity that is given initially in the manifest image of ourselves. The manifest image is in turn embedded in our everyday ways of dealing with the world, each other and ourselves, which include but are certainly not limited to prediction and explanation of each other's behavior. Consequently, clarifying the ontological status of these entities is going to be a matter of insightful observation and description of all those everyday ways of dealing. An appropriate term for this activity is *phenomenology* (Dreyfus 1991 pp.30-39).

A key methodological question, then, is how to conduct phenomenological investigation. Again, this would not be the place to provide an answer, even if I had one. However, it is worth mentioning one point. When trying to clarify the ontology of the manifest image, it is surely permissible to look over one's shoulder at the emerging scientific image. There is no reason to rule that clarification of the manifest image must be done completely independently of knowledge of scientific advances. In particular, when trying to understand what *beliefs* are, there is no reason not to bear in mind any pertinent discoveries from cognitive science. Suppose, then, that Connectionism is indeed painting an accurate picture of the internal states and processes underlying our behavior. And suppose that Ramsey, Stich & Garon are correct that there is nothing in Connectionist models that individual beliefs might be identified with. Then this should count as some evidence that a Dennett-style account of the nature of belief is more likely to be correct. Not conclusive evidence, to be sure; it would have to be weighed in the light of all other kinds of arguments and evidence for the different theories of belief. But on the surface of it, the absence of belief-like entities in Connectionist models of the internal processing should count in favor of a theory of belief that does not require such entities, rather than in favor of

the conclusion that we have no beliefs at all.

(2) The idea that beliefs and desires are inner, causal entities is not something obvious on the surface of the everyday practices in which the manifest image is embedded; rather, it is a particular philosophical interpretation overlaid on those practices. We certainly say things like:

(a) Alice went to the office because she thought she had an appointment.

Some philosophers take a sentence like this to be equivalent to

(b) Alice was caused to go to her office by her belief that she had an appointment.

or

(c) Alice's going to her office was caused by her belief that she had an appointment.

And if her belief is causing Alice to go her office, or her going to her office, what else could that belief be but some discrete mechanism inside her head? Notice however that sentences (b) and (c), and the idea that beliefs are inner causal entities, amount to a particular way of interpreting sentence (a). That sentence does not itself explicitly talk about Alice or her action being caused by anything, or about any inner causal entities.

Once we realize that the idea that beliefs are inner causal entities is a particular interpretation of everyday ways of talking, we can ask how good it is as an interpretation. It turns out to have some serious drawbacks. For example, most people are strongly inclined to *reject* the idea that Alice's belief that she had an appointment caused her to go to her office at least partly because it appears to conflict with the idea that Alice went of her own free will.

Fortunately, there is another theoretical framework for interpreting sentences like (a), one which does not involve the idea that beliefs and desires themselves are causes, and which generally fits much better our everyday practices of understanding ourselves and others in belief/desire terms. That framework, known as the theory of *Agent Causation* (Bishop 1983), holds that the *cause* in any case of human action is not any belief or desire the agent had, but rather the agent herself (hence the theory's name). *Beliefs don't cause actions, people do*, one might say. When an agent acts "because" of a belief, she is said to *take* that belief as a reason for acting, where to take a belief as a reason for acting is to act in such a way that the action makes sense in the light of that belief and everything else the agent does, says and believes. Clearly, on the agent causal conception, there is no requirement that beliefs be inner causal entities; they may perfectly well be the kind of abstract states of a whole system that figure in Dennett's picture.

5. Conclusions

If all this is along the right track, then standard, PDP-style Connectionism does indeed have implications for the mind-body problem, though very different from that proposed by Ramsey, Stich and Garon.

1. Connectionism supports a Dennett-style ontology of belief.

To oversimplify in a useful way, think of the issue this way: prior to Connectionism, there were at least two competing philosophical approaches to the ontology of belief. According to Approach 1, which falls under the Mind as Cognition framework, beliefs are discrete internal causal entities. According to Approach 2, beliefs are abstract, externally attributed states. Connectionism comes along and develops models in which there are no discrete internal entities corresponding to beliefs. Ramsey Stich and Garon, who generally subscribe to Approach 1, take connectionism to imply that there are no such things as beliefs. Surely the more reasonable strategy is to take connectionism as evidence against Approach 1 and in favor of Approach 2.

2. Connectionism throws into question the Mind as Cognition framework.

The view that beliefs are discrete inner causal entities is a core component of the Mind as Cognition framework. More generally, Ramsey, Stich & Garon are working entirely within that general framework. They conceptualize both folk psychology and Cognitive Science as different ways of talking about the inner states, structures and processes that are causally responsible for our behavior. The only kinds of objection that they consider are ones which insist that inner causal entities corresponding to beliefs can in fact be found.

By contrast, Dennett's conception of beliefs is incompatible with the Mind as Cognition picture. On his account, beliefs are not found inside the skull. Since beliefs are surely an essential part of mind, Dennett is following Ryle in throwing into question the overarching framework that unites Hobbes, Descartes, Fodor, Churchland and Ramsey, Stich & Garon.

Connectionism, in supporting a Dennett-style account of the ontology of belief, is therefore throwing into question the whole Mind as Cognition framework. In my opinion, Connectionism implies not that we have no beliefs but that at least three pillars of the Mind as Cognition framework must be rejected.

2.1 Not all of mind is internal. Some aspects of mind, such as beliefs, are much more a matter of what we *do* than internal states or events which are causally responsible for what we do.

2.2 Not all of mind is a matter of the causal underpinning of behavior. To describe an agent as having a belief is a way of figuring out how it will behave without necessarily supposing that the belief is actually some internal mechanism causing that behavior. In one of Dennett's favorite examples, you can predict the behavior of a chess playing program by supposing that it wants to get its queen out early even though there is nothing inside the machine which is plausibly identifiable with that belief.

2.3 Since some aspects of mind clearly *are* internal and a matter of the causal underpinning of our behavior, it follows that mind is not ontologically homogeneous. Mind is made up of a diversity of entities of different ontological categories. Mind itself is not a simple thing of any kind; rather, it is a collection of many different kinds of entities interrelated in complex ways. A useful analogy for mind is an economy. An economy is a totality comprised of a wide range of very different entities - consumers, financial instruments, products, factories, rules, exchanges, etc. Likewise, mind is a complex totality; it might be thought of as an ontological super-category, an entity comprised of many entities of diverse ontological kinds.

3. Mind and Cognition are Ontologically Distinct.

I take cognition, in a very broad sense, to be all the internal states, structures and processes which are causally responsible for our more sophisticated behaviors. Defined this way, cognition includes some of the entities of the manifest image of mind, such as images and processes of conscious silent deliberation. If the above argument is correct, then cognition is just one aspect of the complex totality which is mind. Cognition and mind are not the same thing, any more than factories are identical with economies.

4. Cognitive Science is not the science of the mind; it is the science of cognition, which is just one aspect of mind.

Cognitive Science is the science of cognition. Since mind and cognition are ontologically distinct, Cognitive Science is not the study of the mind as such, but just the study of one critical aspect of mind. Cognitive Science, when it purports to be the science of mind, makes the same mistake as Hobbes, Descartes, Fodor and Churchland - it identifies mind and cognition. Since mind and cognition belong to very different ontological categories, Cognitive Science is making what Ryle would have described as a category mistake. Cognitive Science is no more the science of the mind than a science of factories would be the science of the economy. Thus, an important implication of Connectionism is that there must be a change in the self-image of Cognitive Science. In overthrowing the Mind as Cognition framework, Connectionism is rejecting the idea that Cognitive

Science is the science of the mind.

References

Bishop, J. (1983). Agent Causation. *Mind* 152: 61-79.

Churchland, P.M. (1988). *A Neurocomputational Perspective*. Cambridge MA: MIT Press.

Dennett, D.C. (1987). *The Intentional Stance.* Cambridge MA: Bradford/MIT Press.

Descartes, R. (1911). *The Philosophical Works of Descartes.* Translated by Elizabeth S. Haldane & G.R.T. Ross. Volume 1. Cambridge: Cambridge University Press.

Dreyfus, H. L. (1991) *Being-in-the-World: A Commentary on Heidegger's* Being and Time, *Division 1*. Cambridge MA: MIT Press.

Fodor, J.A. (1987) *Psychosemantics.* Cambridge MA: Bradford/MIT Press.

Gardner, H. (1985) *The Mind's New Science: A History of the Cognitive Revolution*.. New York: Basic Books.

Hobbes, T. (1962) *Leviathan*. New York: Macmillan.

Place, U.T. (1956). Is Consciousness a Brain Process? *British Journal of Psychology*. 44-50.

Ramsey, W., Stich, S.P., & Garon, J. (1991). Connectionism, Eliminativism, and the Future of Folk Psychology. In Ramsey W., Stich S.P., & Rumelhart D.E. (Eds.) *Philosophy and Connectionist Theory.* Hillsdale N.J.: Erlbaum; 199-228.

Ryle G. (1949). *The Concept of Mind.* Chicago: University of Chicago Press.

Sellars, W. (1962). Philosophy and the Scientific Image of Man. In Colodny, R.G. (Ed.) *Frontiers of Science and Philosophy*. Pittsburgh: University of Pittsburgh Press.

1 Actually, the early formulations of the Identity Theory were somewhat vaguely formulated, and if they are reread with a hindsight made sharper by awareness of the distinction between type- and token-identity theses, it can be seen that, while certainly strongly implying a commitment to type-identities, they did not explicitly commit themselves to type-identities. - - - - - -

---

Author: Tim van Gelder

Last updated: 15-Jul-02

Philosophy Department Home Page