Functionalism and Fallibility

John Symons
University of Texas at El Paso

Forthcoming in *Southwest Philosophical Studies*

Functionalism in the philosophy of mind rests on the claim that mental states are multiply realizable; mental states can be realized by or instantiated in a variety of distinct physical structures. To see them as multiply realizable we take mental states as causal roles rather than particular physical structures. As such, functionalism can be contrasted with metaphysical accounts which treat mental states as instances of a mental substance. Instead of puzzling over the relationship between mental and physical kinds, functionalists understand our talk of minds as a way of describing the functions of bodies. The attraction of a functionalist position is obvious. It seems to solve the traditional problem of mind-body dualism, while simultaneously reconciling the apparent irreducibility of psychological discourse with a broadly physicalistic or materialistic ontology.

Functionalism has been challenged insofar as it fails to capture the kind of phenomenal knowledge that results from qualitative experience. While this line of criticism is powerful, I contend that it derives from a more basic weakness of the functionalist position. We can grasp this more basic problem by examining the relationship between functionalism as a metaphysical position and the intuitive assumption that it is possible to engage in progressive inquiry into the nature of minds. At its root, philosophical functionalism requires a stable theory of mind. Inquiry, by contrast, depends on the possibility that we might improve our theories and that, as such, there are features of our current theory of mind which we are likely to abandon.

Functionalism falls into roughly two broad camps, scientific functionalism and analytic functionalism. As we shall see, the former claims that science can, in principle,

provide some finished, true theory of mind and that this theory can be given a functional characterization. Analytic functionalism assumes that we already have a finished, true (though perhaps very minimal) theory of mind. Analytic functionalists assume that folk psychology is true, even though it might consist of a small number of sentences about mental life. According to this view, some set of folk psychological platitudes are taken to be true, can be corralled into a theory and functionalized.

This paper examines the interplay between functionalism as a metaphysical theory of mind and our inquiry into what it is that we believe we are functionalizing. Since inquiry involves assumptions concerning fallibility and since these assumptions are at odds with functionalism, the claim that mental states have a functional character is not as straightforward as is generally believed. Inquiry undermines functionalism's central metaphysical advantage. However, without first engaging in some form of inquiry we can not arrive at a clear idea of the nature of the mind which we are trying to functionalize.

One might argue that failure to account for our need for inquiry or epistemic progress is not a relevant weakness. After all, functionalism is a purely metaphysical doctrine rather than a contribution to philosophy of science or epistemology. However, there are significant risks for the functionalist in distinguishing metaphysical and scientific claims about the mind too sharply. For example, when understood as a purely metaphysical thesis it becomes increasingly difficult to establish why functionalism should count as a theory of mind. As Brie Gertler (1996, 77) and Alvin Goldman (1993, 24) point out, functionalism *per se* neither entails nor explains a distinction between mental and nonmental properties. In one sense, as Gertler and Goldman note, the continuity of the mental and the nonmental is the great virtue of functionalism as a metaphysical position. However, if one is interested in understanding the nature of mind, this apparent virtue is also the source of the potential irrelevance of functionalism, since a theory which does not possess the resources necessary to mark the mental/nonmental distinction can hardly be called a theory of mind.

Functionalism and Psychological Inquiry

In addition to its metaphysical content, functionalism has been connected with concrete empirical hypotheses about psychological states and the prospects for psychology as a scientific theory. It is a small step from a functional characterization of psychological states to a similar treatment of psychological types and laws. A type is functional if all instances of the type are functional and laws can be called functional if they relate some functional type with something else in a law-like way. (See Richardson 1979, 533) Psychological theories can be understood as functional if one agrees that all psychological types are functional and that the laws of psychology relate those types. In this context, functionalist arguments are widely regarded as providing a bulwark against a theoretical reduction of psychology to some more basic or general science. While it is tempting to see the connection between functionalism and anti-reductionism as direct, it requires more argument than is usually acknowledged. For instance, as Robert Richardson shows, reductionism is compatible with functionalism (1979) and so, to be a functionalist is not, automatically equivalent to denying the reducibility of psychological theories. It is important to recognize too that David Lewis' influential presentation of functionalism (1970) was designed to show how psychophysical identifications are possible.

Functionalists tend to combine their arguments for the autonomy of the mental with defenses of favored ways of understanding psychological entities and processes. For example, functionalism is occasionally conflated with representationalism and computationalism. Accepting broadly functionalist metaphysical principles does not, of course, commit one to any particular view of the nature of mental states. However, few functionalists would wish to give up completely on the connection between the metaphysics of mind and our ordinary or even our scientific psychology. Giving up the connection would do more than simply cut off our metaphysics from our science, as we shall see; it would make it difficult to claim that functionalism can provide insight *even* into the metaphysical status of mind. Functionalism would be a barren enough doctrine if it could be shown to make no difference with respect to the conduct of our inquiry but it

would be a real shock if it could be shown to tell us nothing about the mind itself. Stated in its bluntest form, the risk here is that functionalism *per se* can't even tell us whether mental states are multiply realizable?

What then is the status of inquiry for functionalists? This will depend on the type of functionalism we are addressing since different kinds of functionalism will vary with respect to the kinds of inquiry they take to be authoritative. Differences with respect to the preferred method of inquiry will be relevant to our evaluation of different functionalist theories. Commonsense (or analytic or conceptual) functionalists take folk views on psychology as primary. On views of this kind, the way the folk characterizes the causal role of some mental state defines the nature of that mental state; to use a term in a manner contrary to folk usage is to misuse the term. Pain is what pain does and if we don't know what pain's role in the causal economy of nature is, we check with the folks. By contrast, a range of competing forms of functionalism, call them, scientific functionalism for convenience, will see folk accounts as subject to modification in light of evidence from scientific inquiry. However, even when a philosopher defers to the folk in these matters, he will still need to inquire, (unless he takes himself to be an incorrigible authority on all aspects of folk beliefs) in order to determine precisely what it is that the language community has in mind when they use their mental terms. Inquiry takes place in the absence of a complete account of the correct theory of mind.

Identifying mental states

Functionalism (of both the scientific and vernacular varieties) assumes a finished and true theory. On David Lewis' account, for instance, theoretical identifications are made possible by theories rather than being posited independently. This, he argues, "follows from a general hypothesis about the meanings of theoretical terms: that they are definable functionally" (1999, 249) Here, Lewis is following Frank Ramsey's (1929), explanation of how 'theoretical concepts' like *is a proton* or *is a center of gravity* can be treated as existentially bound variables rather than predicates. By 'theoretical concept' he means those which are indispensable for the generalizations and predictions of a theory but whose members are unobservable or otherwise philosophically problematic. Treating

such concepts as existentially bound variables specifies the role of theoretical terms via the system of relationships defined by the structure of the theory. (212-236, 1931) Given some psychological theory, the Ramsey sentence can serve as a way of providing definitions for mental terms that do not themselves include mental terms. By now, the strategy is well known: We begin with some scientific theory T where T ranges over unobservable properties $A_1 \ldots A_n$ and observable properties $O_1 \ldots O_n$. T can be understood as an ordering of observable and unobservable properties; $T(A_1 \ldots A_n , O_1 \ldots O_n)$. In addition to listing properties, it is also possible to introduce reference to individuals $a_1 \ldots a_n$. Now, the ascription of some unobservable property (say the property of being a neutron) to some individual or region of space-time a can be carried out via a sentence containing a higher-order existential quantifier along the following lines:

$$(\exists A_1) \ldots (\exists A_n) [T(A_1 \ldots A_n , O_1 \ldots O_n) \,\&\, A_i a)]$$

This definition characterizes unobservable theoretical terms based solely on existential quantification, observables and the structure provided by the theory. Metaphorically speaking, we can say that they serve to provide non-question begging definitions of mental terms by treating them as locations in the network provided by a theory. If our theory provides a unique ordering of properties, then reference for theoretical terms is fixed via their relationships with one another and with the observable phenomena described by the relevant theory. The structure of relationships between the elements of a theory is presented by the theory and to say that some individual has some property can be converted into a claim about relative placement within the structure described by the theory.

Ramsey elimination does not make any significant difference in the development of a scientific theory of mind since it assumes the existence of a theory that is both finished and true. It tells us nothing about how one might settle on a causal structure appropriate to particular explanations: It assumes an ordering without saying anything about what it is, or how one might decide between alternatives. Of course, Ramsey's account was not originally intended to answer such questions and so this defect does not matter for his purposes. His goal was to account for the meaningfulness of theoretical terms in an established theory. Lewis's use of Ramsey faces the well known threat that even if a part of the folk psychological theory turns out to be false, the Ramseified

version of the theory will also be false. Additionally, as Jaegwon Kim points out, even if the folk psychological theory has false non-mental consequences, the whole Ramsey sentence turns out false (1996, 108)

If we ignore these threats and settle *apriori* on a particular psychological taxonomy and decide that it is not subject to revision, then functionalism suffices as a theory of mind in the sense that it provides a way of resolving the meaningfulness of our talk of mind without encountering ontological worries. This was Lewis' strategy insofar as mental states are "physical states of the brain, definable as occupants of certain folk-psychological causal roles." (1999, 5) By deferring to folk psychology, Lewis' position denies the relevance of progress in psychology to philosophy of mind. This might be a defensible position if it could be shown that we have access to folk psychology in a way which resists correction or refinement via inquiry. It is not clear what it would mean to have access of this kind. Call this kind of access, *access without inquiry*.

By contrast, if inquiry is contributing to the development of improved characterization of mental states via either folk or scientific psychology, our understanding of the explanandum changes. As such the inquiring agent will modify its characterization of the target of scientific explanation. The most minimal kind of change that explananda will undergo over the course of scientific progress results from simply knowing more about the explanandum. Commonsense dictates that we avoid mistakenly latching onto a particular psychological taxonomy prematurely.

Couldn't a scientific functionalist respond, saying that she is not committed to any particular psychological taxonomy, but that she understands mental states in the finished theory as being functionalizable? Insofar as she is invoking a Ramsey sentence, she will be invoking the Ramsey sentence of some true psychological theory. Furthermore, at least for the purposes of metaphysical inquiry, to refer to such a sentence does not require perfect knowledge of it. The trouble for the functionalist is her assumption that mental states as they figure (and if they figure) in the final theory can be identified with the mental states we point to in our current understanding. To claim that the Ramsey sentence will shed light on the mental state that currently interests us involves an unwarranted identification.

Implications for contemporary philosophy of mind

So how do these general considerations play into debates in the philosophy of mind? After all, most philosophically interesting theories of mental states go well beyond the minimal metaphysical doctrine of causal-role functionalism. While they may be couched in the metaphysical framework of functionalism, theories that are worth arguing about are likely to concern the nature of the functions which comprise intelligent behavior and the manner in which they are organized and instantiated. For instance, the generalizations of computational or representational theories might be understood to be true of human minds insofar as they make the claim that there must be particular systems of embodied representations responsible for thought, perception, or action. It is frequently noted that computational functionalism requires an additional non-trivial set of commitments over and above the simple claim that mental states are functional states. For example, the additional assumptions that would turn a minimal functionalist into a computational functionalist might include, for instance, the notion that mental states are representations, that they are interpretable only from within a language of thought, that they are governed by algorithms, etc. Each of the additional computational assumptions is itself a matter for either scientific investigation or arguments from intuition or common sense. Insofar as they go well beyond minimal functionalism, these additional claims do not inherit the metaphysical strengths of that view.

This relatively straightforward point illuminates reductionist criticisms of functionalism of the kind presented by John Bickle. Bickle claims that multiple realizability can be shown to be false in particular cases. For example, the multiple realizability of memory has been shown to be false by the discovery that the molecular mechanisms for memory consolidation are conserved throughout evolutionary history. (2003) Bickle is not arguing that that some other mechanism could not possibly form the basis of memory. His claim is directed specifically towards an empirical reading of Hilary Putnam's original formulation of multiple realizability and concerns the nature of the concept of memory as instantiated in the actual world.

Without addressing the specifics of Bickle's argument, consider whether and where this style of argument confutes minimal functionalism. But, what about multiple realizability? Since one can imagine any particular function being performed by a variety of physical structures, it is possible to think of functions as having a kind of independence from the structures that perform them. In terms of raw logical possibility this is clearly correct. No conceptual barrier blocks the possibility that my current psychological functions could be installed on my computer's hard drive or in a wedge of Swiss cheese. As far as our study of mental life is concerned, Putnam claimed "we could be made out of Swiss cheese and it wouldn't matter." (1975, 291) Given a thin enough account of the concept of thought and a rich enough interpretation of what cheese could be up to, it is logically possible that cheese could think.

Since functionalism is a view concerning the metaphysical status of minds and not a recipe for scientific progress, the strangeness or implausibility of the Swiss cheese mind is not, by itself, a reason to abandon functionalism. For the functionalist, particular realizations of a mental property, whether carbon-, silicon- or dairy-based, are an accidental matter. This is because, according to the functionalist, minds possess no physical features essentially. Functions are simply relative locations within some abstract theoretical structure. If a functionalist believes that minds can only have a causal role via some physical instantiation, then perhaps having *some* physical property is an essential feature of mental entities. However, notice that even this is already a step beyond the core doctrine of functionalism since it draws on a set of beliefs concerning the nature of causal power and entails a physicalist ontology. Minimal functionalism *per se* is not wedded to any one view of what it is to be a cause nor is it committed to a single ontology.

Arguments like Bickle's can only be effective when aimed at a view with more content than minimal functionalism. Without offering some kind of additional, non-functional evidence (empirical, intuitive, or whatever) a minimal functionalist view of memory, for instance, will not go beyond the claim that 'Memory is what memory does'. What we could call the modal refuge defense against empirically-based arguments is powerful but very costly. One must be willing to forego resting one's view on any specific claims concerning what psychological phenomena are in the actual world. That is, the functionalist must detach his metaphysical claim about the multiple realizbility of a

concept from any features of that concept which inquiry might correct. Empirical claims about some mental state must obviously be minimized so as to ensure that one's characterization of the state does not depend on some fact about the actual world that could be shown to be false. This is why the functionalist risks irrelevance.

Functionalists will still wish to say that claims they make concerning, for example, the essential character of memory are applicable, insofar as memory as we find it here in the actual world must also possess the essential properties of memory. However, here the burden falls on functionalists to provide an account of how any proffered list of essential properties for a particular psychological concept is relevant to that concept without thereby being open to correction and without thereby sacrificing functionalism's principal metaphysical advantage.

Conclusion:

When self-described functionalists take sides in disputes concerning the status of folk psychology or the relevance of empirical results, their arguments draw from sources beyond the scope of functionalism proper. Common-sense or analytical functionalism, teleological functionalism, machine functionalism, empirical functionalism… the list of functionalisms is long. However, contention over these positions derives from their prefixes. Arguments from empirical functionalists are, at bottom, empirical; arguments from commonsense or analytical functionalists are derived from the intuitions we claim to have about our own minds, etc. It's the prefixes that allow the functionalist to avoid triviality and to participate in interesting debates.

Whatever merit various forms of functionalism might have in debates over the scientific investigation of mental life, they should not be understood to inherit the strengths or invulnerabilities of the minimal and I would suggest irrelevant "handsome is as handsome does" functionalism. The functionalist qua minimal or non-prefixed functionalist is in no position to help us understand much of anything. In order to support or attack a view on for example the status of folk psychology or the role of neuroscience in philosophy of mind they must resort to non-functionalist premises.

Acknowledgments:

I am grateful to Marc Moffett, Michael Pelczar and Tim Cleveland for their critical comments on an earlier version of this paper. Many thanks to Jorge Valadez for his very helpful response to the paper at the New Mexico/West Texas Philosophical Society Meeting.

Bibliography

Bickle, J. (2003) Philosophy and Neuroscience: A Ruthlessly Reductive Approach. (Kluwer: Dordrecht).

Gertler, B. (1997) "Functionalism's Methodological Predicament". *The Southern Journal of Philosophy* 38: 77-94

Goldman, A. (1993) "The Psychology of Folk Psychology", *Behavioral and Brain Sciences* 16, 15-28.

Kim, J. (1996) *Philosophy of Mind*. (Boulder CO: Westview).

Lewis, D. (1972) 'Psychophysical and Theoretical Identifications', *Australasian Journal of Philosophy* 50, 249-258

Putnam, H. (1975) *Mind, Language, and Reality: Philosophical Papers*, vol. 2 (Cambridge: Cambridge University Press)

Ramsey, F. P. (1931) The *Foundations of Mathematics and other Logical Essays*, ed. R.B. Braithwaite, (London: Routledge & Keegan Paul).

Richardson, R. (1979) "Functionalism and Reductionism" *Philosophy of Science* 46: 533-558.