Functionalism and Sensations

Mark Brown
University of Kansas

Turing Machine Functionalism possesses the rare distinction among philosophical theories of having been definitely refuted. The decisive moment in its history came with the publication of "What Psychological States Are Not"[1] in which Ned Block and Jerry Fodor pressed a number of technical objections against the theory. The Turing machine model of the mind had no place for dispositional predicates such as belief, it was incapable of distinguishing distinct simultaneous mental states and the type identity conditions it placed on mental states were both too fine-grained and too coarse-grained. Functionalism as a general approach to the philosophy of mind survived the demise of Turing Machine Functionalism, but one of the objections urged against the original theory still haunts its descendants. Some mental states, most notably sensations, seem to untutored intuition to have a qualitative character which resists functional explanation. This is not a failure to be taken lightly. The qualitative character of mentality is a pervasive aspect of experience infusing our lives with depth, color and significance. A creature without a qualitative dimension to its existence might be able to behave as we do, but it is hard to see how it could care what happens to it. A philosophy which cannot capture the subjective, phenomenal side of mentality should point beyond itself to a more adequate theory. I will argue that current versions of functionalism fail to do justice to the quality of experience and that they do point beyond themselves to an account which stands a better chance of accommodating both the objective and subjective data.

I

There are many varieties of functionalism currently being defended and criticized, but the central doctrine binding these theories together can be put quite simply: mental states are to be individuated in terms of the abstract causal roles they play. This thesis is as important for what it denies as for what it asserts. It denies that to be in a certain type of mental state

Auslegung Vol. X, No. 3
ISSN: 0733-4311

is to be in any type of physical state. It denies that to be in a certain type of mental state is to be disposed to behave in a certain way. Both the mind-brain identity theory and behaviorism are false on this theory, and for basically the same reason. Mental states are abstract in the sense that they are neutral with regard to the substance of the minds that have them. A mental state need not be realized by a brain state and a mind need not be embedded in a human body. The whole ontological issue of brain, soul or neutral substance is bypassed in favor of a characterization of the conditions which must be satisfied by any possible substance realizing mental properties. This feature of functionalism was inspired largely by the desire to assimilate the insights arising out of recent work in artificial intelligence. What makes the activities of some computers so strikingly similar to human behavior is not the nuts and bolts of the physical machine but the program it executes. Since the same program can be executed by physically disparate computing machines, a certain type of physical structure cannot be essential to its operation. This negative insight of functionalism is its least controversial feature, although we will see reason to doubt it.

On the positive side functionalists typically explicate the notion of an abstract causal role in terms of the causal relations a mental state bears to sensory inputs, behavioral outputs and predecessor and successor states internal to the system. Every mental state is thus implicitly defined in terms of its relations to other mental states and to its internal and external environment. The goal of functionalism can be seen as the elimination of the rough and ready mentalistic vocabulary of common sense psychology in favor of the formally precise language of the kind found in computer programming.

The special difficulty that functionalist theories face with the qualitative character of sensations can now be put plainly. The functionalist program is to define mental states relationally, but I will argue that some mental states appear to have intrinsic properties. The felt phenomenological quality of the rich taste of strawberry shortcake or the feel of skin against fine fur are not plausibly construed as purely relational properties. Furthermore, the qualitative character of a sensation is often crucial to determining the kind of mental state that it is. Could anything be a pain if it felt good, or looked yellow or had no qualitative character at all? At this point the functionalist may well respond that intuitions are easy to manipulate, particularly when the question is framed in such a rhetorical manner. Accordingly, it will be worthwhile to look at a couple of arguments which may provide a firmer guide to intuition.

II

The first argument involves the familiar thought experiment of the inverted spectrum. Suppose it were the case that objective colors (wavelengths reflected by objects) were systematically mapped for different people onto different subjective colors (the way wavelengths appear to perceptual consciousness). If all the similarity relations between subjective colors were preserved,[2] so that red and green remained at opposite ends of the color wheel for example, then it is possible that with respect to color strawberries look to me the way grass looks to you. If the qualitative character of perception is essential to the type identity of perceptual states then you and I are in different mental states when we look at strawberries. Functionalism does not seem to have the resources available for capturing the difference. Our sensory inputs may be the same and our behavioral outputs also might match, and if all the difference and betweenness relations between subjective colors are preserved by the mapping, then my mental state might play the same causal role in the internal economy of my mind as your mental state plays in yours.

The inverted qualia argument has been around at least as long as Locke's Essay[3] and there are a number of standard objections to it. Old-style verificationist theories of meaning have long since been discredited but the feeling lingers on that a proposal upon which no possible evidence could bear must be empty at best. As the situation was described, whenever a person whose spectrum had been inverted saw anything red it would look green to him, but since he has been taught to call things of that color 'red' he would speak in the same way as the rest of us. No one would ever know he was any different. On the other hand, if we do countenance the logical possibility of spectrum inversion an intolerable skepticism about other minds seems to follow. If I cannot know which color my friend sees when he looks at a tomato then how can I know what he feels when he burns his hand? Perhaps what feels like a burn to me feels like an ache or a cut to him. Once the possibility is admitted for colors there seems to be no way to prevent it from undermining our faith in other minds generally.

The temptation to deny the logical possibility of spectrum inversion has often proved irresistable, and subtle arguments have been advanced to show that the idea is incoherent.[4] I will not take the time to examine these arguments because I don't think it is necessary. Instead, consider for a moment the possibility of intra-subjective spectrum inversion. You wake up tomorrow morning and things which looked red yesterday look green to you today. This change would be detectable. You would probably complain loudly

about this unwelcome turn of events. Nor could your complaints be put down to some sort of linguistic confusion since it could easily show up in your behavior as well. You might drive through flashing red lights for example. If it is at least logically possible that this could happen to a person at some time in his life, then what contradiction would be generated by supposing that he has been in this condition his entire life; or that he develops a case of amnesia shortly after his experience of spectum inversion?

A second line of argument open to defenders of functionalism would be to claim that, although the situation described is logically possible, it is nomologically impossible for functionally identical psychological states to differ in their qualitative character.[5] Perhaps there is a mapping which preserves the smooth transitions from one color to another, but the burden of proof would seem to be upon proponents of the argument to show that there is a mapping with the required properties. Although I think that logical possibility is all the inverted qualia argument strictly requires, it is important to recognize that this objection to the argument often expresses the feeling that fanciful thought-experiments carry little weight in the face of an attractive theory. All sorts of extravagant circumstances may be imagined, but in the absence of any relevant empirical data a theory should be judged by its actual successes and failures.

This attitude is understandable but nevertheless misguided. The case is described in the extreme form of fully inverted spectra in order to make vivid the failure of functionalism to capture a salient aspect of our mental life. More modest qualitative disparities are equally good counter-examples to any purely relational theory. Suppose I have a defect in my cornea which imparts a yellowish tinge to my visual field. If a bout with jaundice can produce this condition, why couldn't someone be born with jaundice-like vision? Or perhaps the color wheel that describes your subjective color perception is shifted 2° to the right of the corresponding color wheel of normal human beings. How plausible is the strong thesis that functional identity entails qualitative identity? In all likelihood, descrepancies between subjective quality spaces are actual and rather common.

Functionalists who accept the possibility of inverted qualia are not entirely without resources. Many simply relegate the qualitative character of sensations to some discipline other than psychology and confine their theory to the more cognitive aspects of mentality. I have no quarrel with this strategy, but it is important to realize its cost: the attempt to eliminate mentalistic language from a mature psychology would have to be abandoned. Many cognitive states would have to be partially defined in terms of their

causal relations to unreconstructed sensations, both because sensations often give rise to cognitive states and also because many beliefs, desires and intentions take representations of sensations as their content. Nevertheless, a more modest functionalism is still an attractive and potentially powerful theory. A functionalist theory[6] which defined the sensation of blue, for example, as (a) having some qualitative character (b) systematically situated in a quality space such that (c) its presence reliably signals the presence of objective blue in the environment would be both informative and open to indefinite refinement. The disadvantage of such a theory is, of course, that it leaves the predicate "has qualitative character" undefined.

We are now in a position to provide a partial answer to the skeptical doubts the inverted qualia argument inspires. The first point to notice is that it is not true that no possible evidence could bear upon the truth of this hypothesis. In the case of intrasubjective spectrum inversion the most natural explanation of the situation would be that the neural mechanism which formerly took objective red into subjective red now takes objective red into subjective blue. A fancy brain scan could confirm the hypothesis which then could be used to test for inter-subjective specturm inversion. We will have occasion to return to the suggestion that the realization of a functional theory is the proper level of description for qualitative states, but it may be objected at this point that evidence which is not currently available can hardly explain our confidence in ascribing mental states to other people. Part of the answer to this objection can be found in the functionalist theory just sketched. The fact that other people exhibit similar behavioral responses to color stimulation is evidence for an isomorphism between quality space matrices. Whatever may be the intrinsic character of the sensations of other people, we at least know that their sensations are systematically related in a way which mirrors the relations which hold in our own case. Beyond this, I think that the skeptical doubts are justified. Different people probably do perceive the world differently and the nature of the difference is likely to remain hidden.

III

The second argument is related to the first and can be more briefly stated. If functionalism has no resources for representing qualitative similarity and difference, how can it represent the contrast between a creature whose internal states have no qualitative character at all and the kind of richly qualitative mental life with which we are all familiar? It seems

that the intrinsic features of mental events could slip through a relational net entirely undetected. Imagine a case in which the biochemical equivalents of gears and pulleys determine the transition probabilities from state to state and ultimately overt behavior. Such an individual would be an imitation man, mimicking human life and creating the illusion of mentality where there is none.

This possibility is generally considered to be more damaging to functionalism than the possibility of inverted qualia, and functionalists have been quick to deny its intelligibility. One well-known argument[7] against the possibility of absent qualia can be stated succinctly. In order for alleged cases of absent qualia to be a counter-example to functionalism, qualitative states can have no causal relations to other mental states. Qualitative states do give rise to beliefs about qualitative states, they are often motives for action, etc. Therefore the absence of qualitative states would be detected by an adequate functionalist theory.

This objection misconstrues the point of the original argument. Clearly, qualitative states do play a prominent role in our mental lives, but this role, as the functionalist has defined it, could very well be filled by some other mechanism wholly lacking in qualitative character. Both the absent qualia argument and the inverted qualia argument can be seen as making the same point, i.e., that functionalism neglects some of the causal consequences of the normal operation of the system and some of these consequences may well be relevant to how a specific realization of the functional theory fills the abstract role in question. For example, a functionalist theory of the generation of electricity may well be neutral with respect to whether the electricity is generated by a hydroelectric or by a nuclear process. However, the use of uranium as opposed to water can have dramatic causal consequences which would not be captured by this functionalist account. Likewise, in making the notion of abstract causal role constitutive of mental states, functionalism has pitched its analysis at too abstract a level. A theory of abstract causal roles by its very nature selects from among the causal consequences of a process those consequences which are relevant to the succession of states.

A psychological theory which brought into clear focus the perceptual mechanisms employed by known sentient creatures would nicely complement the abstract functionalism we have been discussing. Fortunately, a thriving model for such a theory can be found in functional explanation in biology. The patterns of explanation found in evolutionary biology suitably adapted to the subject matter of psychology, could supply a much needed element of empirical content to the formal

structure of abstract functionalism. I will now briefly state what I take to be the principal components[8] of functional explanation in biology and also indicate how this explanatory framework can be brought to bear upon the issue at hand.

(1) For the purposes of functional explanation, only causal consequences need be considered, but since our metaphysics presupposes an ontology of individuals, it will be more natural to speak of an item and what it does. The point to notice here is that more than one item can engage in the same activity. If two or more items have the same causal consequences they are said to be functionally equivalent. This aspect of functional explanation in biology is similar to the negative insight of abstract functionalism discussed earlier, although it should be noted that the constraints placed upon functional equivalence are much stronger here.

(2) In system S the causal consequences of an activity must be computed within the context of the containing system, if any. The competing demands of other components in an instantiated system will often force tradeoffs in efficiency. Optimal design for any component of a system cannot be assumed.

(3) Relative to enrironment E the interaction of the system with the outside world must be factored into the derivation of causal consequences. Since a system needs to deploy its resources as efficiently as possible, the demands placed upon it by the environment will shape the configuration of its component parts.

(4) Relative to purpose P not all the consequences of an activity will be functional within the system. The function of the heart, for example, is to circulate the blood, but the activity of the heart also has nonfunctional consequences (heart sounds) and dysfunctional consequences (heart attacks). The introduction of a purpose provides a device for distinguishing between the function of an item and its other causal consequences. The use of teleological language need not involve an appeal to consciousness, entelechies, vital processes or anything of the sort. All that the use of "purpose" entails is that the system is the product of a selection mechanism of some kind. In biology this role is played by natural selection and if natural selection can be reduced without explanatory loss to the operation of mechanistic forces than talk of purposes will be in a sense eliminable. In any case, the use of purpose-laden language is both indispensible as a practical matter and ontologically harmless if relativized to some fully causal selection process.

A functionalism guided by the biological model of explanation will take as its proper object some particular system or kind of system. One would expect naturally evolved systems such as human beings to have roughly the kind of mental states they ought to have

given their epistemic needs, the demands of the environment and the cognitive resources at their disposal. Broadly speaking, the purposes served by sensations are to allow the system to extract information about how the world is impinging upon its body and about the present state of equilibrium in its body. But it is not enough merely to extract the information. The information must be presented to the relevant subsystem in a form in which it can be sorted, processed and used.

Any information flow requires some channel through which it is represented. Information just is the structuring of some channel. Now, the flow of information can be studied from two quite different perspectives. One may concentrate on the amount and type of information to which the system has access and the ways in which the information is utilized and affects the succession of internal states. This is the project which abstract functionalism is best suited to handle. Or, one may focus upon the properties of the channel in which information is embedded. This is the domain of the more concrete functionalism we are now considering. Abstract functionalism will impose some constraints on the character of the channel. The channel must be relatively free from noise and equivocation for example. It must be a clear channel[9] in the sense that the channel either adds no new information or adds only redundant information. Beyond this it is necessary to look at the specific mechanism which carries the information and the broad systemic constraints imposed upon it. One possibility in the case of sensations is to view the qualitative character of sensations as the encoding channel through which the encoded information is presented to consciousness.[10] The distinctive character of each sensory modality could be attributed to the character of the channel in which the relevant information was embedded. The distinctive character of a given sensation would in turn be explained in terms of the encoding principles of the channel in question. In the case of pain the categories of throbbing, cutting, aching and burning together with various levels of intensity and shading between pure types could serve as channel conditions through which information is transmitted. Clearly, much more would need to be said before this suggestion could be fairly evaluated, but being programmatic is not in itself a defect. If a proposal is free from conceptual confusion and promises to uncover heretofore hidden relationships then it is worthy of careful attention.

IV

It may be objected at this point that whatever the virtues of examining the instantiations of functionalist theories, this approach will shed no more light upon the quality of experience than its more abstract cousin. John Searle is one philosopher who is keenly aware of the inability of theories of a functionalist stripe to capture what he calls intrinsic mentality. One of his more recent inventions is the argument from anesthesia.[11] Searle asks us to assume that functionalists have provided a complete specification of the causal role of pain in relation of all possible inputs, outputs and successor states. We may add that the ideal functionalist theory takes due account of the broad systemic constraints imposed by its instantiation in human beings. The theory is recast in the form of a computer program or machine table description and committed to memory by some person. That person is then subjected to anesthesia and asked to mentally rehearse the sequence of states which the theory defines as a searing burning sensation. If he has performed his task flawlessly, Searle claims that our anesthetized person will be an instantiation of an ideal functionalist theory of pain yet fail to have the mental states the theory ascribes to him.

This case is similar to the absent qualia argument, but I think that it is easier to see where this argument goes wrong. A functionalist theory of pain will include a description of states X, Y, Z and the transition probabilities between them. As instantiated, X, Y, Z may correspond to diminished attention span, laying down memory traces and initiation of aversive responses. In contrast our anesthetized friend goes through a series of mental states but not the right ones. The occurant mental states he has are representations which take as their content functionalist descriptions of pain and are not to be identified with the sequence of states to which a functionalist theory is committed. The representational mental states a, b, c, (not equal to) X, Y, Z are wholly unsuited to play the causal role played by pain states. Consequently, the imagined case is no counter-example to a serious functionalism.

Searle recommends insisting upon the first-person perspective when talking to functionalists. It is only from this point of view that the elusive quality of what it is like to be someone can be seen as the challenge to functionalist theories of mind that it is. Searle is absolutely right about this, but for the wrong reasons. To have a first-person perspective of the causal role of a pain state is to be a participant in the causal nexus in which the pain state is produced. In order to know what it is like to be a bat one must first become a bat and this is a transforma-

tion which no mere theory can perform. It is tempting to view the world from the eyes of Berkeley's God, as a wholly representational structure explainable without residue within some theoretical framework. But the world is not a theoretical construct; rather it is an arena of immediate causal interactions, and for creatures with the right kind of internal structure these interactions give rise to immediate qualitative experience. Functionalism may give us some insight into the requisite kind of internal structure, but no theory can ever put us literally into someone else's mind.

NOTES

[1]Ned Block and Jerry Fodor, "What Psychological States Are Not," The Philosophical REview, Vo. 81 (1972).

[2]Cf. Sidney Shoemaker, "The Inverted Spectrum," Journal of Philosophy, Vol. LXXIX, No. 7 (July, 1982).

[3]An Essay Concerning Human Understanding, Peter H. Nidditch, ed. (New York: Oxford 1975), p. 389 (Bk II, chpt. XXXII, sec. 15).

[4]See especially P. F. Strawson's argument that self-ascriptive uses of psychological predicates presuppose other-ascriptive uses in Individuals, Part One, III, 4.

[5]Block and Fodor, op. cit.

[6]For a similar account see Sidney Shoemaker, "Functionalism and Qualia," Philosophical Studies, Vol. 27 (1975).

[7]Ibid. See also Ned Block, "Are Absent Qualia Impossible The Philosophical Review, Vol. LXXXIX, No. 2 (April, 1980).

[8]The account given here is largely derivative from William Wimstatt, "Teleology and the Logical Structure of Function STatements," Studies in History and Philosophy of Science, Vol. 3, No. 7 (1972).

[9]Fred Dretske, Knowledge and the Flow of Information (Cambridge, Mass.: MIT Press, 1981).

[10]For a similar account of secondary qualities see Clifford Hooker, "An Evolutionary Naturalist Realist

Doctrine of Perception and Secondary Qualities," Minnesota Studies in the Philosophy of Science, Vol. IX, 1978.

[11]John Searle, "Analytic Philosophy and Mental Phenomena," Midwest Studies in Philosophy, Vol. VI, 1981.