NARROW FUNCTIONALISM AND BEHAVIOR

Ray Rennard

1. Motivating Narrow Behavior

Individualism in the philosophy of mind is a view about the way that psychological kinds ought to be individuated for the purposes of scientific psychology. A positive statement of individualism is Stephen Stich's *principle of autonomy*: "[T]he states and processes that ought to be of concern to the psychologist are those that supervene on the current, internal, physical state of the organism."[1] According to Stich, psychology has no need for psychological kinds that are individuated in terms of properties that fail to supervene on the current internal state of a subject—in particular, facts about its physical, linguistic, or social environment. Stich offers what many take to be a persuasive argument in defense of the principle of autonomy, called the *replacement argument*:

> Suppose that someone were to succeed in building an exact physical replica of me—a living human body whose current internal physical states at a given moment were identical to mine at that moment. And suppose further that while fast asleep I am kidnapped and replaced by the replica. It would appear that if the crime were properly concealed, no one (apart from my kidnappers and myself) would be the wiser. For the replica, being an exact physical copy, would behave just as I would in all circumstances. Even the replica himself would not suspect that he was an imposter. But now, the argument continues, since psychology is the science which aspires to explain behavior, any states or processes or properties which are not shared by Stich and his identically behaving replica must surely be irrelevant to psychology (165-66).

Stich contends that he and his replica would behave identically *because* they share identical internal physical states, suggesting that only the current internal states of an organism are necessary for psychological explanations of its behavior.[2]

The success of the replacement argument depends crucially on the assumption that Stich and his replica behave identically. *Do* they behave identically? The answer depends on how one construes behavior. On some conceptions of behavior, Stich and his replica behave differently. Suppose that Stich gives his fortune to charity by signing a piece of paper. Stich's replica, were he to have been in Stich's place, would engage in different behavior—viz., an act of forgery. From a legal standpoint, Stich and his replica do not engage in the same behavior since only Stich can legally give away his fortune by signing that paper. (If the kidnapping were ever discovered, Stich would not be responsible for the replica's actions nor would the replica's actions be legally binding.) Let us call behavior that depends on facts about the subject's environment—whether it be facts about the subject's social, linguistic, or physical environment—*wide* behavior. On the wide conception of behavior, Stich would behave differently from his replica in countless other ways. For instance, only Stich can vote legally, fulfill (old) promises, renew his wedding vows, or return home. These are all behaviors that presuppose facts about Stich's history or social environment. Sameness of all

current internal physical properties does not entail sameness of wide behavior in identical contexts.

Stich admits that he and his replica would not engage in the same behavior under every *description*. But not all ways of describing behavior are relevant to psychology. For example, we can describe Jones eating a banana in numerous ways such as eating Aunt Bessie's favorite fruit, eating an item that sells for 49 cents per pound, eating a piece of fruit picked at 2:04 p.m. last Tuesday in Costa Rica, and so forth. What relevance does the fact that the banana was picked at 2:04 pm last Tuesday in Costa Rica hold for explaining Jones eating it? Presumably none. Psychology is not responsible for explaining behavior under every description. Then, how ought behavior be described for psychological purposes? This is a question about the conditions of *individuation* for behavior—the way that behavior is identified, specified, and classified for the purposes of psychology. The conditions of individuation of behavior ought to depend on the explanatory practices of psychology. Stich contends that psychology is only responsible for explaining what he calls "autonomously described behavior"—what I will call *narrow* behavior.

What is narrow behavior? According to Stich, narrow behavior is behavior that satisfies the condition that "if it applies to an organism in a given setting, then it would also apply to any replica of that organism in that setting"(167). But this definition requires independent motivation—especially since it presupposes the very conception of psychology that he claims the replacement argument supports. If the replacement argument is to succeed in supporting individualism in psychology, then it requires (1) an independently specified account of narrow behavior, and (2) an argument to the effect that narrow behavior is the appropriate conception of behavior for psychology. I am skeptical that there is a viable independent notion of narrow behavior. In this paper I focus on the narrow functionalist approach, of which Stich's own account is a version. It will be helpful first to review the case against the behaviorist's conception of behavior as mere bodily movements.

2. Behaviors As Bodily Movements

Clearly, Stich and his replica engage in the same *bodily movements*; perhaps narrow behavior can be equated with bodily movements. Something like this seems to have been favored by the logical behaviorists. According to logical behaviorism, talk of mental phenomena can be *analyzed* in terms of sets of stimulus-response correlations—i.e., as dispositions to engage in a range of behaviors under certain stimulus conditions. Logical behaviorists are not opposed to the *use* of "mentalistic" language in psychology—at least as a first approximation of the kinds of psychological phenomena that need to be explained—so long as it is understood that such talk ultimately must be analyzed into non-mentalistic language.

Behaviorism is often criticized for claiming that mental talk can be analyzed in this way. Consider having the belief that it will rain. According to behaviorism, believing that it will rain is nothing more than being disposed to engage in various behaviors under certain conditions; for example, upon hearing that it will rain, one might carry an umbrella, wear a raincoat, utter, "It will rain," and so on. Forget about the likely fact that the list of possible stimulus conditions and their behavioral manifestations is very large, if not open-ended. The real problem with the behaviorist's account is that being disposed to carry an umbrella is not *sufficient* for having the belief that it will rain (under the cited condition) unless one

also *desires* to stay dry and *believes* that one can stay dry by using an umbrella. Moreover, being disposed to carry an umbrella upon hearing that it will rain is consistent with the attribution of a variety of mental states. One might still believe nonetheless that it will be sunny, and hope to use the umbrella to shade oneself, or one might believe that the umbrella is the reincarnation of one's childhood pet, and desire to take it on a walk. Choosing among the many alternatives requires making assumptions about the individual's other mental states and rationality. I think that these well-documented problems are insuperable for the behaviorist. However, I want to point out something I take to be even more devastating: the behaviorist's conception of *behavior* is unsuitable for psychology.

The term "behavior" is ambiguous. Sometimes we use it to refer to a particular token behavior, such as my raising my hand right now, and sometimes to refer to types of behavior, such as hand-raisings in general. Identifying a token of some type requires the possession of the conditions of individuation for that type. For example, to identify a particular series of bodily movements as a token of the type *hand-raising behavior*, one must have some independent way of identifying that type. Psychology, after all, explains particular phenomena by subsuming them under generalizations that quantify over types, and it is interested in particular behaviors only insofar as they are tokens of behavioral types. So, the behaviorist's claim about behavior must be formulated more carefully. Behavioral types are to be identified (ultimately) with types of bodily movements.

Consider my raising of my arm. There is a strong correlation between arm-raisings (behavior) and arm-risings (bodily movement). There can be no arm-raising without arm-rising. But not all arm-risings are arm-raisings. Suppose my hand rises because a powerful magnet attracts the ring on my finger. Clearly, this does not qualify as an instance of hand-raising behavior. We might rule out such cases by requiring that a bodily movement have an "internal" cause before it can qualify as a bit of behavior. When the magnet lifts my hand, the cause of my hand's rising is external; whereas when I raise my hand the cause is internal. But this cannot be the whole story either, for there are instances of internally caused bodily movements that we would not classify as behavior. Consider spasms. Your hand might rise as the result of random neural activity, but this would not count as behavior on your part. Spasms are things we undergo, things that happen to us. Such events do not call for *psychological* explanation. Behaviors, we like to think, are things we *do*. This suggests that in order for a series of bodily movements to qualify as behavior, it must have the right kind of internal cause.

Even if we concentrate only on the simple cases, the identification of behavioral types with types of bodily movements is problematic. First, being a bodily movement of some particular type is not *sufficient* for being a behavior of some type. Tokens of the same bodily movement type may, on separate occasions, be associated with different behavioral tokens. For example, a certain series of movements of my index finger may be associated with computing a tip on a restaurant bill on one occasion or dialing a friend on the telephone on another. Second, being a bodily movement of some particular type is not *necessary* for being a behavior of some particular type. A behavioral type may have multiple bodily movement realizations. For example, play behavior in humans can be manifested in a wide range of verbal and molar bodily movements—each of which can also be associated with a different behavioral type on another occasion (e.g., fighting behavior).

Nevertheless, we associate token bodily movements with behavioral tokens—for example, we identify certain movements of an individual's hand as an instance of greeting

behavior. What makes the bodily movement qualify as an instance of waving, rather than an instance of dancing or exercising? One plausible (and intuitive) answer is that a token of hand moving counts as an instance of waving behavior if it was caused by the *intention* to wave. We can overcome the failure of bodily movements alone to serve as necessary and sufficient conditions for behavior by appealing to mental states as *causes*. But this approach is not available to the behaviorist. The behaviorist seeks to analyze away all talk of mental states in terms of talk of dispositions to behave under certain stimulus conditions, and takes dispositions to behave to consist of dispositions to engage in bodily movements. However, if a token bodily movement qualifies as an instance of behavior only if it is caused by the appropriate mental state, then the behaviorist cannot analyze away mental talk after all. Behavior cannot be identified with bodily movements alone.

3. Narrow Functionalism.

I now turn to the functionalist response to behaviorism. While functionalism does provide an account of the mind that is better in many ways than behaviorism, it does not, for the most part, offer a better conception of *behavior*. As we have seen, an analysis of a mental state must refer to more than its relations to stimulus conditions and behavior—it must also acknowledge its relations to other mental states. This is one of the central tenets of the functionalist approach. According to functionalism, mental states are defined in terms of their *causal role*—in particular, their casual relations to perceptual stimuli, behavioral responses, and other mental states. This might suggest that the functionalist has no non-circular way of defining mental terms. However, as David Lewis has shown[3] (using a modification of a technique developed by Frank Ramsey), there is a way to eliminate mental terms *en masse* from a functionalist psychological theory, leaving behind only non-mental expressions. On this approach, mental state types can be specified without making explicit reference to other mental states. This is worth considering in some detail.

Consider a psychological theory T that contains a number of mental terms $t_1,\ldots,t_n$. Psychological theory T consists of a conjunction of true statements involving both mental and non-mental terms. The definition of any particular mental term consists in its relations to other mental terms and the non-mental terms in T. Letting S_i be the *i*-th conjunct in T, we get

(1) $T = (S_1 \wedge \ldots \wedge S_n) = \wedge_{1 \leq i \leq n} S_i$.

The psychological theory here is expressed as a conjunction of statements, $S_1,\ldots,S_n$. Formula (2) represents a statement, S_i, with its mental terms explicit.

(2) $S_i[t_j,\ldots,t_k]$

Generalizing to T as a whole, we get:

(3) $T = \wedge_{1 \leq i \leq n} S_i[t_j,\ldots,t_k]$

The right side of the equation expresses T as a conjunction of statements whose mental terms are explicit. Suppose we want to define a particular mental term, t8. Recall that the definition of t8 is given by its relations to the mental and non-mental terms in T as a whole. We can abstract away t8's role in T by applying existential generalization to t8—which consists of replacing every instance of t8 in T with a variable, and existentially quantifying over that variable to create a definite description. We can define t8 as follows:

(4) $t_8 =_{def} (\iota x) \wedge_{1 \leq i \leq n} S_i[t_j, x, \ldots,t_k]$

(Roughly, t_8 is the unique thing that satisfies such and such a role in T.) This gives us a definition of t_8, but one that still makes explicit reference to the other mental terms in T. However, there is a way to specify T devoid of all explicit mention of its mental terms. We simply replace each instance of every mental term type with a variable bound by an existential quantifier. What we get is called the "Ramsification" of T:

(5) $T^R = (\exists x)\ldots(\exists z) \wedge_{1 \leq i \leq n} S_i[x,\ldots,z]$

Notice that T^R contains no mental terms, but only non-mental terms and relations as well as logical notation. Mental terms are represented by variables, whose values are determined relative to their role in the theory as a whole. Finally, we can define any particular mental term in T by converting its existential quantifier into a definite description. For example, (6) gives the definition of mental term t_8:

(6) $t_8 =_{def} (\iota v)(\exists x)\ldots(\exists z) \wedge_{1 \leq i \leq n} S_i[v, x,\ldots,z]$

Remember that t_8 entered the theory T by occurring in causal generalizations connecting its referent to perceptual stimuli, behavioral responses, and other mental states. When T is Ramseified, t_8 and all the other mental terms are replaced by bound variables. Accordingly, the definition of t_8 in T^R contains no mental terms. It is important to note that functionalists are realists about the mind. After all, mental terms are replaced by *existentially* bound variables. However, aside from the logical vocabulary, all that remains in the definition of a mental term in the Ramseified theory are the non-mental terms of T—viz., the perceptual stimuli and behavioral responses (and logico-mathematical expressions) mentioned in T. So, there is a sense in which mental states are *reduced* to stimulus-response correlations, without being *eliminated*.

So what does all of this say about the functionalist's account of behavior? Well, if the functionalist is to provide a non-circular definition of mental states in terms of non-mental states via the Ramsey-Lewis method, then perceptual stimuli and behavioral responses must be specified entirely in non-mental terms. How the functionalist characterizes behavioral responses (and perceptual stimuli) depends on the choice of the underlying psychological theory. There are two main options. The underlying theory is either a common sense psychological theory or a scientific one. Following Ned Block, we can distinguish between two forms of functionalism (271-2). Conceptual functionalists take the underlying theory to be common sense psychology, while Psychofunctionalists assume that the underlying theory is an empirical one. As Block notes, the Conceptual Functionalist is committed to the inputs and outputs as individuated by the folk—viz., perceptions of objects in the environment and widely described behavior, respectively. In my discussion of behaviorism, I suggested that a token of some bodily movement type qualifies as a token of some behavioral type if it was caused by the appropriate (content-laden) mental states. This does appear to be the conception of behavior to which folk psychology appeals. The behaviorist could not make this move because it requires positing the existence of mental states that the behaviorist hoped to define away by appealing to behavior. It might seem that the Conceptual Functionalist is in a better position because he acknowledges the existence of mental states as causes of

behavior. But not so. If he wants to reduce mental talk to non-mental expressions via the Ramsey-Lewis method, then a token behavioral response cannot be characterized as a token bodily movement with a particular mental state cause. The expressions that remain in the Ramseified psychological theory cannot be individuated *themselves* in terms of their mental state causes.

The Psychofunctionalist, on the other hand, is free to characterize the inputs and outputs however it is congenial to the underlying empirical psychological theory. The Psychofunctionalist will typically characterize the inputs and outputs in neurophysiological terms—e.g., in terms of patterns of afferent and efferent neural activity. However, such a "narrow" specification of inputs and outputs leaves the Psychofunctionalist in no better position to characterize behavior than the Conceptual Functionalist since it merely draws the line at the nervous system. We can run the same argument on the efferent neural firings as we did on the molar bodily movements. Suppose that some pattern *P* of neural activity normally results in a particular type of hand and arm movement. On what grounds is the Psychofunctionalist to identify patterns of type *P* with waving behavior rather than fly-shooing behavior, if both behavioral types are consistent with *P*? The natural response is to argue that an instance of neural activity of type *P* qualifies as an instance of waving behavior if it was caused by a certain combination of mental states, which themselves are identified in part by their causal relations to other mental states and patterns of afferent neural activity. But this response is not available to the Psychofunctionalist who wishes to reduce mental state talk to non-mental vocabulary using the Ramsey-Lewis method, since distinguishing between identical patterns of neural activity that can be associated with different molar behaviors requires appealing in part to mental state causes.

4. The Syntactic Theory of Mind

Stich's own preferred solution—the Syntactic Theory of Mind (STM)—is a version of Psychofunctionalism which is described as:

> The basic idea of the STM is that the cognitive states whose interaction is (in part) responsible for behavior can be systematically mapped to abstract syntactic objects in such a way that causal interactions among cognitive states, as well as causal links with stimuli and behavioral events, can be described in terms of the syntactic properties and relations of the abstract objects to which the cognitive states are mapped.[4]

STM is a form of Psychofunctionalism because (a) cognitive state types are characterized as syntactic types which are defined in terms of the causal-functional roles their tokens play in the specified system, and (b) what role a syntactic type has in a system is a matter for empirical (mainly, computational) theorizing in cognitive psychology. What distinguishes STM from other forms of Psychofunctionalism is the absence of appeals to the *contents* of cognitive states in the specification of cognitive phenomena.

While STM need not be committed to Ramsey-Lewis reductionism, it still faces a problem with individuating sensory inputs and behavioral outputs in a way that is jointly narrow, syntactic, and appropriate for psychology. A syntactic type is defined in terms of the relations between its tokens and tokens of other types within the system. A syntactic system is not required to be narrow, in the sense of being defined over tokens "within the skin" of an

organism. A syntactic description of the summing of a list of numbers can appeal to syntactic tokens within and without the skin of the adder—e.g., the manipulation of tokens in an external memory store (say, a scratch pad).[5] However, if the syntactic types are narrow—as the replacement argument requires—then they cannot involve relations to external factors. Narrow STM must, at best, advert to syntactic characterizations of *narrow* properties—such as bodily movements or neural activity. And, again, it is far from clear that such abstractions of narrow properties are appropriate for psychology. First, in virtue of what are two tokens of some type-identical narrow property, such as a series of bodily movements or efferent neural activity, to be distinguished as different syntactic types? For example, if two tokens of the same series of bodily movements or sequence of neural activity are associated with fighting on one occasion and playing behavior on another, how are they to be distinguished as tokens of different syntactic types according to narrow STM? An intuitive response is that they can be distinguished on the basis of their causal relations to perceptions of foe and friend respectively. But this raises the question, now with respect to perception. How are afferent neural patterns of foe and friend to be distinguished? Second, in virtue of what are tokens of different narrow types to be classed as tokens of the same syntactic type? Fighting behavior can be manifested by an organism through a multitude of different bodily movements or series of efferent neural activity, and the narrow facts alone are insufficient to bring together this gerrymandered set under one syntactic type. The difficulty is only compounded, moreover, by considering differences across organisms within a species, not to mention differences among organisms of different species.

A strong case can be made that psychology does not characterize behavior narrowly. Cognitive psychology is in the business of characterizing the main cognitive capacities—including perception and behavior—and explaining how they are realized in organisms. A cognitive capacity is characterized in terms of its function—what it is *supposed* to do for the possessing organism. For example, the function of binocular vision is to enable the possessing organism to distinguish correctly the relative depth-relations among objects in its visual landscape. The capacity for binocular vision can fail—e.g., in cases of illusion such as those involved in the Ames room—suggesting that its function cannot be identified with what it is *disposed* to do. The function of binocular vision is to discern depth-relations correctly even if it sometimes (or mostly) does not. Cognitive psychologists are interested for the most part in understanding the physical mechanisms that realize these capacities, and there are two ways to construe this project. One may see it as an exercise in uncovering the underlying physiological processes (what makes them tick), or as an attempt to understand why they work as they do (why they tick, as it were). There is nothing particularly *psychological* about the former project, it seems to me. There, appealing to the function of a cognitive capacity plays merely a heuristic role, enabling the researcher to focus on one set of physical properties rather than another. The latter project, on the other hand, seems eminently psychological. But here, the function served by the physical mechanism takes precedence: we are interested in the physical mechanism only insofar as it enables the organism to possess a particular cognitive capacity.

The function of a cognitive capacity, I claim, cannot be understood by appealing to narrow facts alone; it is necessary to advert as well to environmental and historical factors. Two type-identical physical mechanisms may serve diverse cognitive functions, given differences in environmental or historical facts. Appealing to these differences, one can distinguish them as tokens of different cognitive types. Furthermore, radically different

physical mechanisms can be typed together as tokens of the same cognitive type, given relevant similarities in environmental or historical facts. Applied to behavior, we class together tokens of different bodily movement (or neurophysiological) types in virtue of their common intentional properties or goals—neither of which can be recovered from their narrow specifications.

I have argued that if the replacement argument is to provide support for individualism in psychology, it requires a viable notion of narrow behavior. Neither behaviorism nor Psychofunctionalism—including STM—succeeds in providing a suitable notion of narrow behavior. Consequently, as a motivation for individualism in psychology, the replacement argument fails.

NOTES

[1] Stich, *Folk* 164. Cf. Jaegwon Kim's Explanatory Thesis: "Internal psychological states are the only psychological states that psychological theory needs to invoke in explaining human behavior—the only states needed for psychology" (183).

[2] Fodor presents a somewhat similar argument, which goes as follows. If two people engage in different behavior, the causal powers of their internal states must differ. Since Stich and his replica are, *ex hypothesi*, type-identical internally, their internal states cannot differ in their causal powers. Therefore, Stich and his replica must engage in the same behavior.

[3] See "Terms" and "Identifications."

[4] Stich 149. Also see Stich (1991), where he makes it clear that mental state types are to be individuated with respect to sensory inputs and behavioral outputs, specified syntactically. He calls this "fat syntax."

[5] See Clark for similar examples. The very possibility of defining a computational system over tokens outside an individual is presupposed by the so-called "systems reply" to John Searle's Chinese Room thought experiment.

WORKS CITED

Block, N. "Troubles with Functionalism." *Perception and Cognition. Issues in the Foundations of Psychology, Minnesota Studies in the Philosophy of Science* 9. Ed. C. W. Savage. Minneapolis: U of Minnesota P, 1978. Rpt. of N. Block, ed., *Readings in the Philosophy of Psychology* 1. Cambridge, MA: Harvard UP, 1980.

Clark, A. *Microcognition*. Cambridge, MA: MIT P, 1989.

Fodor, J. *Psychosemantics*. Cambridge, MA: MIT P, 1987.

Kim, J. "Psychophysical Supervenience.". *Supervenience and Mind*. Ed J. Kim. Cambridge: Cambridge UP, 1982.

Lewis, D. "How to Define Theoretical Terms." *The Journal of Philosophy* 67 (1970): 427-46. Rpt. of D. Lewis. *Philosophical Papers*, 1. New York: Oxford UP, 1983.

———. "Psychophysical and Theoretical Identifications." *The Australasian Journal of Philosophy* 50 (1972): 249-58. Rpt. of D. Lewis, *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge UP, 1999.

Stich, S. *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT P, 1983.

———. "Narrow Content Meets Fat Syntax." B. Loewer and G. Rey, eds., *Meaning in Mind: Fodor and his Critics*. Oxford: Blackwell, 1991.