SRI International

A COGNITIVIST REPLY TO BEHAVIORISM

Technical Note 348

March 1985

By: Robert C. Moore
Artificial Intelligence Center
Computer Science and Technology Division

SRI IR&D Project 500KZ

This note is a commentary on "Behaviorism at Fifty" by B. F. Skinner, and appears in The Behavioral and Brain Sciences, Vol. 7, No. 4, pp. 637-639, December 1984.

Preparation of this paper was made possible by a gift from the System Development Foundation as part of a coordinated research effort with the Center for the Study of Language and Information, Stanford University.

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

ABSTRACT

The objections to mentalistic psychology raised by Skinner in "Behaviorism at Fifty" (Skinner, 1984) are reviewed, and it is argued that a "cognitivist" perspective offers a way of constructing mentalistic theories that overcome these objections.

There are two major themes running through Skinner's various objections to mentalistic psychology. He argues, first, that mentalistic notions have no explanatory value ("The objection is not that these things are mental, but that they offer no real explanation..."), and second, that since the correct explanation of behavior is in terms of stimuli and responses, mentalistic accounts of behavior must be either false or translatable into behavioristic terms ("...behavior which seemed to be the product of mental activity could be explained in other ways."). What I hope to show is that a "cognitivist" perspective offers a way of constructing mentalistic psychological theories that circumvent both kinds of objection.

The first theme appears twice in infinite-regress arguments. Skinner ridicules psychological theories that seem to appeal to homunculi, on the grounds that explaining the behavior of one homunculus would require a second homunclus, and so on. Later he employs the same rationale to criticize theories of perception based on internal representation: If seeing consists of constructing an internal representation of the thing seen, the internal representation would then apparently require an inner eye to look at it, etc. Skinner's concern for explanatory value is also evident in his view of mental states as mere "way stations" in unfinished causal accounts of behavior. If an act is said to have been caused by a certain mental state, without any account as to how that state itself was caused, there seems to be little to constrain what states we invoke to explain behavior. The limiting

case would be to "explain" every action an agent performs by simply postulating a primitive desire to perform that action.

Skinner's concerns about explanatory value should not be taken lightly, and they seem to me to pose serious problems for older-style mentalistic psychological theories. Often these theories appear to allow no direct evidence for the existence of many kinds of mental states and events. According to such theories, "poking around the brain" will not help, because mental entities are not physical; moreover, asking the subject for introspective reports may not help either, because mental entities can be unconscious. But a second consequence of the view that mental entities are nonphysical is that we have no a priori idea as to what the constraints on their causal powers might be. We are thus left in a situation in which we could, at least in principle, postulate any mental states and events we like, adjusting our assumptions regarding their effects on behavior to fit any possible evidence.

How does cognitivism avoid Skinner's charges in this area? I take it that what distinguishes cognitivism from other mentalistic approaches to psychology is the premise that mental states can be identified with computational states. This has two consequences for the problem at hand. First, computational states must in some way be embodied in physical states. This means that if behavioral evidence alone were not sufficient to determine what mental state an organism was in, neurological evidence could be brought to bear to decide the question.

Second, and of much more immediate practical consequence, is the fact that there is a very well-developed mathematical theory of the abilities and limits of computational systems. Hence, once we identify mental states with computational states, we are not free to endow them with arbitrary causal powers.

When a computational account of mental states and events is given, Skinner's infinite-regress arguments lose their force. While it is a characteristic of computational theories of mind to explain the behavior of the whole organism in terms of interactions among systems that may appear to be "homunculi," a computational account, as Dennett (1978, pp. 123-124) has pointed out, requires each of these homunculi to be less intelligent than the whole they comprise. Thus, while there is indeed a regress, it is not an infinite one, because eventually we get down to a level of homunculi so stupid that they can be clearly seen to be "mere machines." Similar comments apply to Skinner's worries about explaining perception in terms of mental representation. Although he is quite correct in maintaining the pointlessness of supposing that the brain contains an isomorphic copy of the image on the retina, computational theories of vision simply do not work that way. Although they make use of internal representations, these express an interpretation of the image, not a copy. While a retinal image might be thought of as a two-dimensional array of light intensities, the postulated representations take as primitives such notions as "convex edge," "concave edge," and "occluding edge." These representations are

then manipulated computationally in ways that make sense given their interpretations. Waltz (1975) gives a very clear (albeit already outdated) exposition of this approach.

Skinner's notion of unfinished causal account is not necessarily answered simply by adopting a computational perspective, but conscientious cognitive theorists do address the problems raised by the tendency to attribute precisely those structures that are needed to account for observed behavior. Some deal with it as Skinner suggests, by investigating the causation of mental states (e.g. studying language acquisition), but the more frequent strategy is to show how a single computational mechanism (or the interaction of a few mechanisms) accounts for a broad range of behavior. If, for example, we can show that a relatively small set of linguistic rules can account for a much larger (perhaps infinite) set of natural-language sentence patterns, then it is certainly not vacuous, or without explanatory value, to claim that those linguistic rules in some sense characterize the mental state of a competent language user.

Whether or not Skinner would acknowledge that the cognitivist framework has the potential to produce mentalistic theories with genuine explanatory value, I suspect he would argue that, because of the other major theme of his paper, any such conclusion is really beside the point. In his view, mentalistic terminology is at best a rather complicated and misleading way of talking about behavior and behavioral dispositions. Skinner's picture seems to be that mental states, rather

than being real entities that mediate between stimulus and response, are merely summaries of stimulus-response relationships. Thus, hunger, rather than being what causes us to eat when presented with food, would be regarded as the disposition to eat when presented with food. (This interpretation of mental states obviously reinforces Skinner's opinion that mental explanations of behavior are vacuous; attributing eating to a disposition to eat explains nothing.)

The response to this point of view is that, even if we could get a complete description of an organism's "mental state" in terms of behavioral dispositions, that fact would not vitiate attempts to give a causal account of those dispositions in a way that might make reference to mental states more realistically construed. A computer analogy is helpful here. Complex computer systems often have "users' manuals" that are intended, in effect, to be complete accounts of the systems' behavioral dispositions. That is, they undertake to describe for any input (stimulus) what the output (response) of the system would be. But no one would suppose that to know the content of the user's manual is to know everything about a system; we might not know anything at all about how the system achieves the behavior described in the manual. Skinner's response might be that, if we want to know how the behavioral dispositions of an organism are produced, we have to look to neurobiology--but this would miss the point of one of the most important substantive claims of cognitivism. Just as in a complex computer system there are levels of abstraction above the level of electronic components

(the analogue, one supposes, of neurons) that comprise coherent domains of discourse in which causal explanations of behavior can be couched ("The system computes square roots by Newton's method."), so too in human psychology there seem to be similar levels of abstraction--inlcuding levels that invlove structures corresponding roughly to such pretheoretical mentalistic concepts as belief, desire, and intention.

Finally, it may very well be impossible to describe the behavioral dispositions of organisms as complex as human beings without reference to internal states. Skinner seems to assume uncritically that, if the sole objective of psychology is to describe the stimulus-response behavior of organisms, one can always do so without reference to internal states. But this is mathematically impossible for many of the formal models one might want to use to describe human behavior. In particular, given some of the behavioral repertoires that human beings are capable of acquiring (e.g., proving theorems in mathematics, understanding the well-formed expressions of a natural language), it seems likely that no formal model significantly less powerful than a general-purpose computer (Turing machine) could account for the richness of human behavior. In a very strong sense, however, it is generally impossible to characterize the behavior of a Turing machine without referring to its internal states. Now, the behaviorists may be fortunate, and it may turn out that the behavioral dispositions of humans are indeed describable without reference to internal states, but Skinner appears not even to realize that this is a problem.

To summarize: (1) Skinner's arguments against the explanatory value of mentalistic psychology do not apply to properly constructed cognitivist theories; (2) the existence of a complete behavioristic psychology would neither supplant nor render superfluous a causal cognitivist account of psychology; (3) the regularities of human behavior that Skinner's approach to psychology attempts to describe may not even be expressible without reference to internal states.

REFERENCES

Dennett, D. C. (1978) Brainstorms (Bradford Books, Publishers, Montgomery, Vermont).

Skinner, B. F. (1984) "Behaviorism at Fifty," The Behavioral and Brain Sciences, Vol. 7, No. 4, pp. 615-621, December 1984.

Waltz, D. L. (1975) "Understanding Line Drawings of Scenes with Shadows," in The Psychology of Computer Vision, P. Winston, ed., pp. 19-91 (McGraw-Hill Book Company, New York, New York).