Some Limitations of Behaviorist and Computational Models of Mind

John Collier
Department of Philosophy
University of Calgary
January 10, 1986

ABSTRACT

The purpose of this paper is to describe some limitations on scientific behaviorist and computational models of the mind. These limitations stem from the inability of either model to account for the integration of experience and behavior. Behaviorism fails to give an adequate account of felt experience, whereas the computational model cannot account for the integration of our behavior with the world. Both approaches attempt to deal with their limitations by denying that the domain outside their limits is a part of psychology. These attempts to turn the shortcomings of the two models into virtues would be more convincing if their limitations were not diametrically opposed. I will argue that in each case the limitations are too restrictive unless the theories are augmented by physiology.

Behaviorism is either false or trivial unless it can distinguish between cognitive and non-cognitive activity. Analogously, cognitivism is either false or trivial if it cannot distinguish between representational and non-representational activity. To avoid this, both behaviorism and cognitivism need to be augmented either with physiology, or else with each other. I assume, since most of us have only a rudimentary knowledge of physiology, that "folk psychology" intuitively follows the second route by integrating rudimentary versions of behaviorism and cognitivism. A scientific theory would have to make this intuitive integration explicit. Given the mutual under-determination of behavioral functions and cognitive algorithms, it seems that the only way to explicate the relations between the two is in terms of their regular causal interactions. Barring some sort of supernatural medium, this will inevitably involve physiological theory.

Introduction

The purpose of this paper is to describe some limitations on scientific behaviorist and computational models of the mind. These limitations stem from the inability of either model to account for the integration of experience and behavior. Behaviorism fails to give an adequate account of felt experience, whereas the computational model cannot account for the integration of our behavior with the world.

Both approaches attempt to deal with their limitations by denying that the domain outside their limits is a part of psychology. Behaviorism denies that the internal activity of the mind is relevant to scientific investigation of psychology, while cognitivists believe the mind must be understood as if there were no external world (Fodor 1980, Pylyshyn 1980). These attempts to turn the short-comings of the two models into virtues would be more convincing if their limitations were not diametrically opposed. I will argue that in each case the limitations are too restrictive unless the theories are augmented by physiology.

Behaviorism

Scientific behaviorism holds that the psychology of an animal (or class of animals) can be determined by observing the regular correlations between environmental stimuli and consequent behavior. Inasmuch as these correlations are regular, the animal can be treated as a black box. Irregular relations between stimuli and responses are of no interest to psychology, since they
cannot be studied scientifically. The regularities required can be found either in individual animals, or across a group of similar animals. Consequently, there is no reason why correlations which are necessarily singular for a particular animal (as might be involved in one-trial learning) cannot be studied.

Objections to behaviorism, aside from Chomsky's (1967) objections involving speed of language learning (which some behaviorists, such as Quine (1969, p 95) find unobjectionable) and various moral objections which seem scientifically irrelevant, stem largely from the feeling that something has been left out; the internal activity of the animal must be important.

Nelson (1969) has attacked behaviorism on the grounds that for any pair of finite strings of inputs and outputs there is an arbitrarily large number of functions which can produce the particular output string from the given input string. Consequently, scientific behaviorism is in principle incapable of determining the function correlating stimuli to responses, since complete determination would require an infinite string of stimuli.

This is no doubt true if no upper limit can be placed on the information processing capacity of the "black box". However, there seems to be nothing in principle against bringing non-psychological constraints into consideration in applying psychological theories, just as there is nothing wrong in considering physical or psychological constraints on biology. Physical or biological miracles may occur in animals with mental capacities, but there is no strong reason to believe this, and good reason to doubt it. The very size and construction of the brain places some upper limit on its information handling capacity. This upper limit may be too large to make behaviorism of practical value to the working psychologist, but it does undermine the claim that behaviorism is in principle false. More practically, evolutionary biology gives us good reasons to believe that whatever functions describing an animal's capacities at birth are relatively simple, since they had to evolve from simpler capacities. Developmental psychology allows us to extend this consideration recursively.

Chomsky (1967) has objected that there is no non-trivial way to distinguish stimuli from non-stimuli and responses from activity which is not psychologically relevant, or else that the behaviorist tautologically builds in to his theory of the animal that certain inputs and outputs represent stimuli and responses. This objection is a bit too quick. We can distinguish stimuli and responses through their falling under the appropriate law-like relations, such as the law of effect. Although it might be objected that these laws, which constitute the psychological theory, thereby define the subject matter to be explained, other sciences, such as physics, work in the same way (see Kuhn (1970), Sneed (1971), and Stegmuller (1976)), so the objection holds against all theoretical science. The real problem is that various purely organic responses of the animal also obey the supposedly psychological laws, and behaviorism has no principled way to eliminate them. Some examples are immunity, drug tolerance, and perhaps even tanning.

Unless we are prepared to hold that there is nothing in principle distinguishing psychology from physiology, we need a principle for distinguishing behavior from merely organic activity. What distinguishes purely organic responses, it seems, is that the animal has no direct awareness of them. Behaviorists need to distinguish between cognitive and non-cognitive activity. Although they have attempted to do this (Skinner 1974, Natsoulas 1980), there are problems.

First, it is conceivable that there are distinct psychological states which are not behaviorally distinguishable. For example, consider an actress who is so good at acting that no matter what situation she is in she never gives herself away. She behaves the same way as someone who is not acting, but it seems quite plausible that she could nonetheless know that she was acting. She might, for example, have a strong subjective experience of the falsity of her persona, but be so exquisitely guarded that it is never, and can never, be revealed in her behavior.
Behaviorists might argue that such purely subjective differences, which have no effect on behavior, are psychologically irrelevant. This argument begs the question, and conflicts with common sense. Despite the undeniable methodological difficulties in studying non-behaviorally manifested psychological differences, they should not be ruled out a priori.

The behaviorist might instead argue that the mere possibility of psychological differences which are not, at least potentially, manifested in behavior does not falsify behaviorism. Behaviorism, it might be claimed, is an empirical theory, and should be accepted as long as it can account for the available evidence, which does not presently suggest there are different psychological states which can not be discriminated on behavioral grounds.

Even given this reply, it seems that behaviorism fails as an adequate explanatory theory. Behaviorism may be descriptively adequate (the objection goes), but it lacks explanatory adequacy (Chomsky, 1965). Specifically, although behaviorism may be able to give nomically necessary and sufficient conditions for any particular psychological state, it provides us with little insight into the subjective relations between states. Even behaviorists tend to impute intentional states to the animals they study, and when their language is purged of such imputations, their explanations of behavior seem very spare indeed. It seems that an explanatorily adequate psychology must not only accurately discriminate psychological states and describe their relations, but must also accurately represent the way the subject coordinates its psychological states, inasmuch as it does so.

Skinner attempts to account for cognitive activity in terms of private stimuli and responses. Natsoulas correctly rejects Skinner's "repeated translations of mental episodes" in terms of private stimuli and responses as "alien to behaviorism" (1980, p 150). In developing his own theory, Natsoulas explicitly assumes mind-brain identity (1980, p 152), though he need only augment behaviorism with a theory of the physiology of consciousness. There does not seem to be another behaviorist alternative.

**Computationalism**

The computational view of mind focuses directly on internal experience. It takes into consideration the internal activities of the animal, and bases the theory of mind on them alone. Since not all internal activities are cognitive, the psychologically relevant activities are those which are "cognitively penetrable", i.e., alterable by purely cognitive factors such as goals, beliefs, inferences, and so on, and the fixed functional capacities (functional architecture) of which the cognitively penetrable activities are composed (Pylyshyn 1980). The argument for ignoring external factors such as stimuli and responses is based on the underdetermination of the interpretation of mental representations by the mental activity itself. Just as the function from stimulus to response underdetermines the algorithm used to compute the function, which function a given mental activity is performing depends on the interpretation of the function, and is not fully specified by the mental operations alone. We could imagine two computers following exactly the same algorithm, but connected differently to input and output devices, which could therefore compute different functions. Since, by assumption, the mental activity is what is psychologically relevant, the function computed is outside of the domain of psychology. The proper domain for psychology is the set of laws governing the computations of the mind, i.e., the algorithms the mind executes.

Demopoulos (1980) argues that if underdetermination of function is an argument for considering the interpretation of mental representations to be irrelevant, then if it can be shown that the algorithm computed is similarly underdetermined, the motivation for the distinction in relevance is undermined. He goes on to argue that the finiteness of the input/output data and the internal computational steps do not fully determine the algorithm computed, since there are a variety of algorithms compatible with any finite set of actual computations.
This argument is similar to Nelson's argument against behaviorism. Pylyshyn (1980) replies that the underdetermination is just a particular instance of the general problem of induction. He admits that there may be a problem concerning the determinacy of learned rules, since these are presumably represented, and are subject to the same indeterminacy as any other representations. He points out, though, that this difficulty does not arise for algorithms directly executed by the functional architecture of the mind itself. Inasmuch as the learned algorithms are composed out of the functional architecture, they are also determinate.

Since representational activity is so central to psychology, any adequate psychological theory should explain how cognitive activity is representational, and should allow us to differentiate representational activity from non-representational activity. Cognitivism has a problem with representational activity which is analogous to the problem behaviorism has with consciousness. If any activity of the functional architecture is necessarily representational, then all cognitive activity is representational, but this seems to be false. Common sense tells us that some cognitive activity is purely manipulative, involving only the processing of representations. We need a non-trivial basis for distinguishing representational activity. The most natural basis is an appropriate connection to, or the capability of being appropriately connected to, input and output, whether these are understood as afferent and efferent processes or as stimuli and responses.

Suppose this were not true, and that representational activity could be distinguished from non-representational activity solely by its internal functional role. If so, given that not all cognitive activity is representational, there would a functional distinction between representational and non-representational activity. That is, the representational would itself have to be represented. This representation would itself have to be functional, and so on. This generates an infinite regress which is vicious, since every step would have to be expressed functionally. It seems that either all cognitive activity is representational, which is false, or else criteria external to cognitive activity, namely input and output, are involved in distinguishing the representational from the non-representational.

Conclusions

The behaviorist risks trivializing his theory unless he can give a principled distinction between conscious (cognitive) behavior and non-conscious activity. Following Natsoulas, this either requires augmenting behaviorism with some sort of cognitivism, or else a knowledge of the activity of the brain. Likewise, the cognitivist risks trivializing his theory by making all mental activity representational, including activity which we would normally say only manipulates representations, unless he can distinguish between representational and non-representational activity. He cannot do this internally, on pain of vicious infinite regress. Either he must take into consideration non-cognitive connections of cognitive activity to afferent and efferent signals, which invokes physiology, or else he must invoke regular connections to stimuli and responses.

Both behaviorism and cognitivism need to be augmented either with physiology, or else with each other. I assume, since most of us have only a rudimentary knowledge of physiology, that "folk psychology" intuitively follows the second route by integrating rudimentary versions of behaviorism and cognitivism. A scientific theory would have to make this intuitive integration explicit. Given the mutual under-determination of behavioral functions and cognitive algorithms, it seems that the only way to explicate the relations between the two is in terms of their regular causal interactions. Barring some sort of supernatural medium, this will inevitably involve physiological theory.
References


