

A BEHAVIORIST'S DEFINITION OF CONSCIOUSNESS ¹

BY EDWARD CHACE TOLMAN

University of California

This paper should have been called "The frantic attempt of a behaviorist to define consciousness." In fact, the doctrine I shall present seems even to me quite unprovable and to you it will no doubt seem something far worse. And yet so great is my faith that behaviorism must ultimately triumph that I should rather present even the following quite doubtful hypothesis than hold my mouth and say nothing. If we behaviorists can not present good theories, we can at least present as many bad ones as possible in order that by their successive refutation we may be forced finally either into discovering the correct theory, or, if there be none, into abandoning our behavioristic adventure altogether.

Before, however, attempting my definition of consciousness, let me first briefly outline the nature of behavior as I see it. Every behavior-act, in so far as its continued going-off is contingent upon there proving to be such and such specific features in the environment, must be said in so far to postulate or cognize those features. For example, when a rat, after learning, is ready to enter only the white alley of a discrimination box, and not the black alley, the continuance of this 'white-entering' behavior must be said to express a cognitive postulation as to the difference between white and black. Further, the fact that the continuance of this tendency to enter the white and not the black is also contingent upon the further circumstance that there prove to be a difference between food and non-food on the two sides of the box must be said to express also a cognitive differentiation between food and non-food. And, thirdly, the continuance of this entering

¹ Read before the Western Psychological Association, Los Angeles, Calif., June 17, 1927.

of the white rather than the black must be said to express in addition a cognitive differentiation as to the relative sign-relationships of white and black—the one as indicative of the food and the other as indicative of the non-food. For, if any one of these three sets of environmental facts should suddenly change, this behavior-act would break down. Thus, if there ceased to be a difference between the white and the black, or between the food and the non-food, or between the respective sign-relations of the white and the black with regard to food and non-food, the consistent entering of the white and avoidance of the black would no longer continue. In short, the continued going-off of this white-entering behavior-act assumes, postulates, these three specific sets of environmental facts and relations.

But this which we have thus proved for the discrimination-box behavior holds, it would seem, in a similar manner for all behaviors. Every behavior-act, in going-off and being what it is, expresses, implies, certain specific characters in the environment. And this is so because the continuance of its going-off can be shown to be contingent upon there actually proving to be such characters in the environment. If these expected characters are not found, the act sooner or later ceases or modifies itself. Behavior is driven by organic needs, and in going-off it postulates that the environmental characters and relations are such that it will prove an appropriate behavior for satisfying those needs. The going-off of a particular act postulates a particular complementary character in the environment. And this is to be assumed as true for all behavior-acts whether new and just learned or old and well-established by custom. The only condition is that these acts show themselves ready for alteration, if things go wrong.

In spite, however, of this conclusion that practically all behaviors are thus cognitive or postulative, we must note the further fact that many of these cognizing and postulating behaviors are none the less quite obviously automatic and unconscious. For a behavior to be postulative and cognitive, it is not necessary that it also be conscious. A well-established and quite automatically functioning habit act cognizes and

postulates the environment, in our sense that its continuance is contingent upon the environment proving actually to be so and so. But such a habit-act may none the less be quite unconscious.

What, then, we must now ask, is the further occasion and cause of consciousness? Our answer will be that *wherever an organism at a given moment of stimulation shifts then and there from being ready to respond in some relatively less differentiated way to being ready to respond in some relatively more differentiated way, there is consciousness*. For example, let us assume that our rat has up to some given occasion been responding in undifferentiated fashion to the white and black alleys. He may, perhaps, have been treating them both as distinct from some third chromatically colored alley, but as between these two, the white and black themselves, his behavior has not distinguished. On this particular occasion, however, we assume that something internal happens, such that he then and there switches from not being ready to respond to them as differentiated to being ready to respond to them as differentiated. The moment of this switch is the moment of consciousness. The organism then and there becomes conscious of the difference between black and white. On all the previous occasions his behavior treated the black and white as alike. On the occasion when the switch occurs his behavior first starts to treat them as different. It is this change to the new differentiation which we define as consciousness. The behavior after such a switch may in time become just as automatic as the behavior before it. Acts which imply more cognitive differentiation may be just as automatic as ones which imply less cognitive differentiation. It is only the switch-over when it occurs in a given moment of stimulation that defines consciousness.

What, now, is the mechanism of such switch-overs? In order to answer, we shall have to consider a new principle. This new principle is that organisms, at least the higher ones, are to be assumed capable not only of actual behaviors but also of what may be called mere behavior-adjustments. The nature of these behavior-adjustments is to be assumed such

that they in some manner bring the animal into contact with the same stimulus-results with which he would be brought in contact, if he should actually behave. The results of any proposed act may thus by means of a mere feint or adjustment to that act be brought into the present and become a conditioner for or against the act. To make an adjustment to an act is to achieve a *representation* (based, of course, upon what has happened upon previous occasions when this act or similar ones have actually been performed) of the probable stimulus-results to be expected from the act.

This doctrine of an ability by virtue of mere behavior-adjustments to represent the probable results of acts may well strike you as a pretty mystical affair. It may sound to you unworthy of consideration by any hard-headed scientist, let alone a behaviorist. And yet, I would ask you, what is Watson's own doctrine of implicit or sub-vocal speech, in so far as it has any cogency, other than a specific account of just such behavior-adjustments? Watson, of course, does not call them behavior-adjustments, but his gestures and sub-vocal contractions seem to have been devised by him to serve the very function which we are ascribing to the behavior-adjustment. His doctrine has cogency only in so far as he implies that gestures and sub-vocal speech serve to bring (*i.e.*, represent) to an acting or listening organism the type of stimulus results to be expected from an actual overt behavior, if it were carried out. The baby, when ruminating on what it wants, says sub-vocally among other things the word 'doll.' But this saying of 'doll' serves to represent the type of stimuli to be expected if the baby were actually to go and get the doll. If these represented doll stimuli are satisfactory, the child performs the actual act of going and getting. If not satisfactory, he rehearses, sub-vocally perhaps, the names of other toys.

The above is, of course, not quite the way Watson's own argument reads. It is, however I believe, the way it should read, and it is only because of such an implied reading that his argument has such cogency as it seems to have. I present this account to you here, however, not to ask you to accept

the doctrine of sub-vocal speech and sub-gesture as such, but rather to woo your minds gently to my own more general notion of the behavior-adjustment. The behavior-adjustment, whatever its neurological or physiological character, is to be conceived functionally as a surrogate for actual behavior,—and a surrogate which somehow serves to bring into the present, that is to make then and there active upon the organism, the stimulus-results to be expected from the corresponding actual behavior.

Assuming for the purposes of argument that you accept this doctrine, the next step will be to declare that it is these behavior-adjustments which produce or *are* consciousness. When a rat on some given occasion switches over from a condition of non-readiness to discriminate white and black to one of readiness to discriminate them, and, as we have said, thereby becomes conscious of the difference between them, this switch-over and this consciousness are mediated, we shall now declare, by a behavior-adjustment. In this case we shall assume it is a behavior-adjustment to the act of running and looking rapidly from the one color to the other. The stimulus-results which would come from such an actual running or looking would presumably be a complex pattern containing both the run-from color and the run-to color. The rapid passage from the one to the other would, that is, presumably result in a sort of Gestalt (?) containing both the colors set off in juxtaposition one against the other. And the adjustment to such a running would be to bring this resultant Gestalt into the moment before actual behavior. Thus, it would be possible for the animal, when faced with either color alone, to respond nevertheless discriminatively to the difference between them. Another point, however, must now be noted, namely, that after this new differentiating behavior has once become established, consciousness and the behavior-adjustment can apparently drop out and yet the new discriminating behavior continue.

We must assume that the complex stimulus-pattern of white in juxtaposition to black, or vice versa, is still needed for the continuation of the discriminating behavior. But we

shall assume that eventually this complex pattern results automatically by pure associative extension from the white stimulus alone or the black stimulus alone. Simple redintegrative bonds must become established whereby the stimulus-results from running back and forth are now automatically fused into either the white or the black stimuli alone.

So much for the consciousness of white versus black. We saw, however, at the beginning of this paper that the total behavior of choosing one alley rather than the other involves not only this differentiation of white from black, but also a differentiation of food from non-food. And it involves likewise a differentiation of the specific sign-connection of white from that of black. We now assert that the initial appearance of these other two differentiations also involves consciousness. And they also are to be explained by the functioning of behavior-adjustments.

The switch-over at some single moment of stimulation from not being ready to differentiate between food and non-food to being ready to differentiate between them would be mediated by a behavior-adjustment for running rapidly from the one goal to the other. Such a behavior-adjustment would present the complex Gestalt-result of the two types of goal compared one against the other. And the mediating presence of such a Gestalt would constitute a then and there consciousness of the food or the non-food character of the particular goal presented or represented.

Finally, the switch-over (on any given occasion) from not being ready to treat the sign-relationship of the black and the white as different to being ready to treat them as different would also be due to the mediating function of behavior-adjustments. In this case the behavior-adjustments would be those for actually going down the presented alley and reaching the to-be-expected food or non-food result. Thereby a fused Gestalt would be produced in which the presented stimulus, white or black, not only would be set over against its comparison color but also would be enlarged by its to-be-expected food or non-food result. Only on the basis of this total Gestalt would the behavior of entering or not entering ensue.

To recapitulate, we would suppose the rat's total process of learning to be something as follows: First, the animal, after a greater or smaller number of trials, would come, when faced with the white or the black, to make an adjustment for running back and forth. And he would thereby become conscious of the whiteness or blackness. Similarly, when faced with the food or the non-food, he would make an adjustment for running back and forth and thus become conscious of the foodness or non-foodness. Finally, when faced with the white or the black, he would also make an adjustment to entering that one or the other and thereby become conscious of the to-be-expected food or non-food result. On the basis of all three such adjustments, resulting, let us say, in one grand total Gestalt, he would respond. This total Gestalt would contain the differentiation of white from black, of food from non-food, and of the sign-relationship of white as leading to food from that of black as leading to non-food. And on the occasions of its first appearances there would be consciousness. On later occasions this grand total Gestalt would eventually come by mere associative extension, without the intervention of behavior-adjustments, *i.e.*, without consciousness.

One last word. You will perhaps be doubtful that the lowly rat is capable of all this. So I think am I. The important point is merely that if a rat learns consciously, the above gives a perfectly objective definition of how he might do it. It may be that he learns unconsciously.² If he learns unconsciously, then we should have to assume that the changes from the readiness for undifferentiated behavior (*i.e.*, behavior mediated by very simply 'Gestalted' stimuli) to the readiness for differentiated behavior (*i.e.*, behavior mediated by more complexly 'Gestalted' stimuli) occurs somehow automatically between trials. We could then assume no mediating adjustments to introduce these changes. We should be forced to suppose that the initial stimuli somehow grow large and properly 'Gestalted' by mere mechanical accretion.

² Though the recently reported results of McDougall and his son (*J. Comp. Psychol.*, 1927, 7, 145-176) tend to minimize the probability of such unconscious learning.