

Two Quotes from Chomsky's Review of B.F. Skinner's *Verbal Behavior*

Behaviorism in psychology was a highly influential movement, of which Skinner was one of the main proponents. It aimed to describe the behavior of animals (and people) in terms of notions such as *stimulus*, *response*, *reinforcement* and *deprivation*. Many of Skinner's experiments were on animals in controlled situations: for example, rats seeking food in a maze, or pigeons pressing bars in response to stimuli such as flashes of light, in order to get food. As such, behaviorism can perhaps be taken as the most serious attempt to reduce psychology to the study of the "instinctive" behavior that Nietzsche and Yoav allude to.

The book *Verbal Behavior* was a substantial attempt to show that the ideas and principles drawn from these studies could be extended to the use of language. Chomsky's review was strongly critical of the claims in the book. I feel that the book and its review are tremendously relevant to the study of AI: behaviorism was quite successful in describing low-level, instinctual behavior; Skinner was one of its best known proponents; *Verbal Behavior* was a serious attempt to extend the results to "higher mental processes"; Chomsky's critique of the book is incredibly thorough, and many of the points he made seem broadly relevant. Indeed, in the preface to a later reprint of the review, Chomsky makes it quite clear that his article was not merely intended as a book review, but was a broader critique of the idea that the results from behaviorism could be extended or scaled to more complex mental processes:

I had intended this review not specifically as a criticism of Skinner's speculations regarding language, but rather as a more general critique of behaviorist (I would now prefer to say "empiricist") speculation as to the nature of higher mental processes. My reason for discussing Skinner's book in such detail was that it was the most careful and thoroughgoing presentation of such speculations, an evaluation that I feel is still accurate. Therefore, if the conclusions I attempted to substantiate in the review are correct, as I believe they are, then Skinner's work can be regarded as, in effect, a *reductio ad absurdum* of behaviorist assumptions. My personal view is that it is a definite merit, not a defect, of Skinner's work that it can be used for this purpose, and it was for this reason that I tried to deal with it fairly exhaustively.

The review is a long, slow read, but is full of great points, so I'd urge anyone interested in AI to read it. I'll just pull out a few paragraphs that I thought were relevant.

[1] Chomsky makes the point that what is remarkable about Skinner's work is not the basic assumptions about the need to model input-output behavior (functional analysis) – this is uncontroversial – but the simplicity of the models that Skinner posits to model this behavior. Chomsky claims that Skinner has underestimated the complexity of the problem, and in particular the complexity of the organism being studied (for evidence of this, see point 2):

It is important to see clearly just what it is in Skinner's program and claims that makes them appear so bold and remarkable. It is not primarily the fact that he has set functional analysis as his problem, or that he limits himself to study of *observables*, i.e., input-output relations. What is so surprising is the particular limitations he has imposed on the way in which the observables of behavior are to be studied, and, above all, the particularly simple nature of the *function* which, he claims, describes the causation of behavior. One would naturally expect that prediction of the behavior of a complex organism (or machine) would require, in addition to information about external stimulation, knowledge of the internal structure of the organism, the ways in which it processes input information and organizes its own behavior. These characteristics of the organism are in general a complicated product of inborn structure, the genetically determined course of maturation, and past experience. Insofar as independent neurophysiological evidence is not available, it is obvious that inferences concerning the structure of the organism are based on observation of behavior and outside events. Nevertheless, one's estimate of the relative importance of external factors and internal structure in the determination of behavior will have an important effect on the direction of research on linguistic (or any other) behavior, and on the kinds of analogies from animal behavior studies that will be considered relevant or suggestive.

Putting it differently, anyone who sets himself the problem of analyzing the causation of behavior will (in the absence of independent neurophysiological evidence) concern himself with the only data

available, namely the record of inputs to the organism and the organism's present response, and will try to describe the function specifying the response in terms of the history of inputs. This is nothing more than the definition of his problem. There are no possible grounds for argument here, if one accepts the problem as legitimate, though Skinner has often advanced and defended this definition of a problem as if it were a thesis which other investigators reject. The differences that arise between those who affirm and those who deny the importance of the specific "contribution of the organism" to learning and performance concern the particular character and complexity of this function, and the kinds of observations and research necessary for arriving at a precise specification of it. If the contribution of the organism is complex, the only hope of predicting behavior even in a gross way will be through a very indirect program of research that begins by studying the detailed character of the behavior itself and the particular capacities of the organism involved.

Later, in the conclusions, he goes even further with this point:

Anyone who seriously approaches the study of linguistic behavior, whether linguist, psychologist, or philosopher, must quickly become aware of the enormous difficulty of stating a problem which will define the area of his investigations, and which will not be either completely trivial or hopelessly beyond the range of present-day understanding and technique. In selecting functional analysis as his problem, Skinner has set himself a task of the latter type. ...

[2] As a central part of the criticism, Chomsky points out that although the behaviorist terminology (*stimulus, response* etc.) is well-defined in the laboratory experiments, it loses objectivity when being applied outside the laboratory, to more complex behavior (i.e., language in this case). Chomsky accuses Skinner of the following sleight of hand:

... [Skinner] utilizes the experimental results as evidence for the scientific character of his system of behavior, and analogic guesses (formulated in terms of a metaphoric extension of the technical vocabulary of the laboratory) as evidence for its scope. This creates the illusion of a rigorous scientific theory with a very broad scope, although in fact the terms used in the description of real-life and of laboratory behavior may be mere homonyms, with at most a vague similarity of meaning. To substantiate this evaluation, a critical account of his book must show that with a literal reading (where the terms of the descriptive system have something like the technical meanings given in Skinner's definitions) the book covers almost no aspect of linguistic behavior, and that with a metaphoric reading, it is no more scientific than the traditional approaches to this subject matter, and rarely as clear and careful.

In the main substance of the review, Chomsky goes on to show this by pulling apart various terms that Skinner uses. Here is part of his critique of the use of the term *stimulus* (similar arguments are given for *response, reinforcement* and so on):

Consider first Skinner's use of the notions *stimulus* and *response*. In *Behavior of Organisms* (9) he commits himself to the narrow definitions for these terms. A part of the environment and a part of behavior are called *stimulus* (eliciting, discriminated, or reinforcing) and *response*, respectively, only if they are lawfully related; that is, if the *dynamic laws* relating them show smooth and reproducible curves. Evidently, stimuli and responses, so defined, have not been shown to figure very widely in ordinary human behavior. We can, in the face of presently available evidence, continue to maintain the lawfulness of the relation between stimulus and response only by depriving them of their objective character. A typical example of *stimulus control* for Skinner would be the response to a piece of music with the utterance *Mozart* or to a painting with the response *Dutch*. These responses are asserted to be "under the control of extremely subtle properties" of the physical object or event (108). Suppose instead of saying *Dutch* we had said *Clashes with the wallpaper, I thought you liked abstract work, Never saw it before, Tilted, Hanging too low, Beautiful, Hideous, Remember our camping trip last summer?*, or whatever else might come into our minds when looking at a picture (in Skinnerian translation, whatever other responses exist in sufficient strength). Skinner could only say that each of these responses is under

the control of some other stimulus property of the physical object. If we look at a red chair and say *red*, the response is under the control of the stimulus *redness*; if we say *chair*, it is under the control of the collection of properties (for Skinner, the object) *chairness* (110), and similarly for any other response. This device is as simple as it is empty. Since properties are free for the asking (we have as many of them as we have nonsynonymous descriptive expressions in our language, whatever this means exactly), we can account for a wide class of responses in terms of Skinnerian functional analysis by identifying the *controlling stimuli*. But the word *stimulus* has lost all objectivity in this usage. Stimuli are no longer part of the outside physical world; they are driven back into the organism. We identify the stimulus when we hear the response. It is clear from such examples, which abound, that the talk of *stimulus control* simply disguises a complete retreat to mentalistic psychology. We cannot predict verbal behavior in terms of the stimuli in the speaker's environment, since we do not know what the current stimuli are until he responds. Furthermore, since we cannot control the property of a physical object to which an individual will respond, except in highly artificial cases, Skinner's claim that his system, as opposed to the traditional one, permits the practical control of verbal behavior is quite false.

Other examples of *stimulus control* merely add to the general mystification. Thus, a proper noun is held to be a response "under the control of a specific person or thing" (as controlling stimulus, 113). I have often used the words *Eisenhower* and *Moscow*, which I presume are proper nouns if anything is, but have never been *stimulated* by the corresponding objects. How can this fact be made compatible with this definition? Suppose that I use the name of a friend who is not present. Is this an instance of a proper noun under the control of the friend as stimulus? Elsewhere it is asserted that a stimulus controls a response in the sense that presence of the stimulus increases the probability of the response. But it is obviously untrue that the probability that a speaker will produce a full name is increased when its bearer faces the speaker. Furthermore, how can one's own name be a proper noun in this sense?

I take the following points from this.

- Language use – and other higher mental functions – are significantly more complex than the “instinctual” behavior that was studied in the original behaviorist experiments. In particular, the simple functions used to describe behavior in the laboratory experiments are unlikely to scale to the study of language, or to any number of higher mental processes, for that matter.
- We should be careful about the generalization of terminology that is well-defined in simple, controlled experiments, but becomes vacuous or tautological when applied to more complex scenarios. See the comment about Skinner's sleight of hand above.
- The results of experiments concerning low-level, instinctive behavior may not easily generalize to explain more complex behavior. Rather than searching for general principles which explain all of intelligence, we should isolate and study manageable problems concerning different types of information processing – speech and language, vision, motor control and so on. No doubt there will be some principles which apply across many of these domains – general results in supervised or reinforcement learning for example – but we should also expect to need specialized machinery for different tasks (e.g., signal processing for speech and vision, similar machinery for vision, bayes networks for representation and reasoning under uncertainty, or the formal machinery of grammars and logics for language).

Mike C.