On Books

Conservative Behaviorism: A Review of Staddon's *The New Behaviorism: Mind, Mechanism, and Society*

G. E. Zuriff
Wheaton College

The “new behaviorism,” as Staddon envisions it, occupies a narrow conceptual space, hemmed in by Skinner's radical behaviorism on one side and cognitivism on the other. Accordingly, his important book, *The New Behaviorism: Mind, Mechanism, and Society*, is a navigational chart on sailing between the Scylla of Skinnerism and the Charybdis of modern cognitive theory. After a review of the history of behaviorism and an introduction to the experimental foundations of radical behaviorism, Staddon launches into a critique of radical behaviorism on empirical, social policy, and theoretical grounds.

EMPIRICAL DIFFERENCES

Staddon uses Skinner's “superstition” experiment to illustrate the empirical flaws in Skinner's concept of the operant. When a hungry pigeon received grain presentations every 12 s, independent of its behavior, in a short while, the pigeon began to show vigorous, stereotyped activity. Skinner's explanation was simple. The delivery of food will inevitably have followed some behavior and will therefore accidentally reinforce it. This behavior, now occurring at an increased frequency, will be highly likely to be occurring again when the next reinforcement is

delivered, and the accidentally reinforced behavior comes to dominate the pigeon's activities.

In challenging Skinner's explanation, Staddon first questions the assumption that the mere contiguity of response and reinforcement is sufficient for the strengthening of a response. Based on experimental results, many his own, he argues that the strengthening of a response requires both contiguity and competition from other responses. That is, the reinforcer strengthens the response it follows only if the response is a better predictor of the occurrence of that reinforcer relative to other responses. This notion of relative predictiveness (“competition”) also calls into question whether Skinner's central concept of *reinforcement contingency* is sufficiently complex to account for the experimental data.

Second, Staddon questions the assumption that happenstance determines which response is adventitiously reinforced. Based on research on autoshaping, instinctive drift, and his own seminal experiments on superstition, Staddon discusses the behavioral laws that determine which responses are strengthened by noncontingent reinforcement. In particular, he demonstrates that reinforcement has both eliciting and selective (strengthening) effects. The superstition experiment can thus be seen as an example of the complex interaction of operant and classical conditioning procedures. Whereas the classical procedures generate a rep-

Staddon, J. E. R. (2001). *The new behaviorism: Mind, mechanism, and society*. Philadelphia: Psychology Press.

Correspondence should be sent to the author at Wheaton College, Norton, Massachusetts 02766 (E-mail: gzuriff@wheatonma.edu).

ertoire of activities to the conditioned stimulus, the operant procedures work to select the strengthened response. Thus, not only is Skinner's simple explanation insufficient but his sharp distinction between operant and respondent is also blurred.

Although these criticisms of Skinner's empirical conception of operant conditioning are familiar, Staddon does integrate them in an informative way, choosing a Darwinian framework. Reinforcement is seen, through an evolutionary metaphor, as a three-stage process. First, varieties of responses are generated through elicitation and instinct. These response variants compete with one another in terms of their causal relations with consequences. Finally, through a process of selection, certain variants are selected and survive. Skinner is faulted for ignoring the processes of variation (how novel responses are generated) and competition (how causal dependency is registered) while concentrating exclusively on selection.

DIFFERENCES IN SOCIAL POLICY

Having rejected Skinner's explanation of the superstition experiment, Staddon, not surprisingly, also rejects Skinner's extrapolated explanation of human superstitions. Indeed, throughout his book, Staddon opposes all of Skinner's attempts to extrapolate explanations from simple experiments with animals to the complexity of everyday human life, derisively (and punningly) terming them "premature extrapolations." In general, Staddon's approach to human psychology is much more cautious than Skinner's. Not only is Staddon less willing to extrapolate, but he is also less willing to assume Skinner's radical revolutionary stance in opposition to conventional societal values, practices, and traditions. In short, Staddon's new behaviorism is a behaviorism for the social and scientific conservative.

Continuing along with the Darwini-

an metaphor, Staddon adopts Skinner's evolutionary epistemology, but he adds his own conservative twist. According to this epistemology, knowledge is a product of our evolutionary history. Beliefs compete, and, through a process of natural selection, those that work best survive to become what we regard as truth. Skinner used this conception of truth for revolutionary purposes. Because behavioral science informs us what works best in the social world, the principles of behavior are a better guide to social truth than religion, tradition, and intuition.

Whereas Skinner thus saw evolutionary epistemology as the justification for overthrowing traditional social values, Staddon draws the opposite conclusions. Staddon notes that we can determine values on the basis of what is more effective for our survival only if we can foresee the future to know what we will need to survive. However, because perfect foreknowledge is not possible, survival of the culture is not a practical guide in most of the hard cases. It follows that successful cultures will have adopted practices whose benefits in the competition with other cultures could not have been foreseen when they were adopted. Therefore, concludes Staddon, a culture with some unquestioned, culturally inherited beliefs is fitter in the Darwinian sense than a culture with fewer beliefs, although there is, of course, no way to know specifically which traditional beliefs will be beneficial. Furthermore, he argues that when traditional values are undermined, it is more likely that society will succumb to values that have not passed the tests of time, many of which will turn out to be harmful. Thus, Staddon's version of evolutionary epistemology suggests a rational basis for faith and tradition in general, although not in the specifics.

Among the cherished social values Skinner urges society to reject are freedom, autonomy, and responsibility. He argues that all three are based upon the illusion of free will. Furthermore,

Skinner argues, these illusory values interfere with the use of behavioral technology to control social behavior for human benefit because people oppose behavioral control, fearing it threatens these values. Hence, many of Skinner's later writings were attempts to move society beyond those illusions.

Staddon, again taking a more conservative approach, challenges Skinner's claims and defends the traditional values. First, Staddon argues that freedom is a subjective feeling whereas determinism, or what he calls "predictability," is an external objective description. Therefore, he concludes, the two are compatible—one may feel free although one's behavior is predictable.

Similarly, Staddon attacks Skinner's claim that determinism precludes personal responsibility. Staddon adopts a utilitarian perspective and argues that the purpose of social reward and punishment is to modify human behavior in socially desirable directions. However, this system can work only if rewards and punishments have their intended and predictable effects. Thus, on the utilitarian view, both our notion of personal responsibility and our judiciary system must assume determinism and would be useless if people truly possessed free will.

On empirical grounds, Staddon also questions another of the basic principles of Skinnerian social policy—the preference for positive reinforcement over punishment. Reviewing the empirical literature, Staddon concludes, as have most previous reviewers, that punishment is as effective as positive reinforcement in modifying behavior, contrary to Skinner's claim. In conjunction with his utilitarian perspective on the judicial system, this conclusion naturally leads Staddon to support the use of punishment to combat crime. In fact, allowing himself some premature extrapolation, Staddon speculates as to whether the increase in lawlessness since 1970 may be the result of the relaxation of penalties over the past 50 years. Once again, Staddon, sounding like a law-and-order conservative, op-

poses Skinner's revolutionary program, arguing that Skinner's proposals to transform the penal system are not compelling.

At a more fundamental level, Staddon challenges Skinner's basic view that behavioral technology should be used to engineer a better society. On empirical grounds, Staddon notes the limits of behavioral control. The effectiveness of contingencies of reinforcement can be diminished by interference from elicited effects of reinforcement, the organism's capacity limitations, the paucity of behavioral variants from which the contingency can select, and a host of other variables.

On conceptual grounds, Staddon argues that a behavioral science cannot provide the goals for a behavioral technology. He rejects Skinner's solution to this problem, namely Skinner's identification of the good with what is reinforcing or ultimately what promotes survival. Staddon argues that Skinner does not tell us what reinforcement is, and the survival criterion presupposes perfect knowledge of the future. Thus, Skinner's solution does not supply a useful guide to the goals of behavior control. But underlying Staddon's intellectual opposition to Skinner lies the soul of a conservative. Staddon simply distrusts Skinner's revolutionary proposal to transform humankind through behavioral science. Or as Staddon states it, "There is always something dangerous in a philosophy that is discontented with human beings as they are and wishes them improved according to some model" (p. 102).

DIFFERENCES IN THEORY

Skinner's radical behaviorism generated its own peculiar philosophy on the nature of scientific theory and explanation. For Skinner, a theory is a parsimonious set of functional relationships serving the purpose of prediction and control. Within a behavioral theory, the most important functional relationships are those between behavior

and contingencies of reinforcement. A good theory will not refer to unseen inferred entities, but will be limited to stimuli and responses, including those occurring inside the body.

Staddon's theoretical behaviorism rejects nearly all of this. He contends that mere prediction and control are insufficient for explanation, understanding, and scientific progress. In his reading of the historical record, the laborious construction of a theory of fundamental principles has yielded, in the long run, greater understanding, scientific breakthroughs, and useful prediction and control than "premature exploitation of a half-baked technology" (p. 150). But what kind of theory is Staddon advocating instead?

Staddon does not favor a theory that explains behavior in terms of contingencies of reinforcement. Although a contingency states what the consequence of a response will be, it does not say whether the response will in fact occur. Hence it can be said to provide a teleological explanation for behavior (i.e., what consequences are achieved by the behavior), rather than a causal explanation (what antecedents brought about the behavior). Staddon, following the Darwinian metaphor, prefers the latter kind of explanation because it takes into account the variation from which the reinforcement contingency selects.

Because we cannot yet specify these proximal causes of behavior, Staddon advocates a "theoretical behaviorism" in which antecedent internal states are inferred, contrary to Skinner's prohibitions. For Staddon, an internal state represents the cumulative effects of the organism's behavioral history, and this state determines how the organism will respond to future stimuli and contingencies. Nevertheless, the state is not identical to this history, because more than one history could have placed the organism in that exact same state with regard to future behavior. Thus, an internal state represents the cumulative effects of all behavioral histories whose effects with regard to future be-

havior are equivalent. Organisms with quite different histories might be in the same internal state in that predictions of their future behavior will be the same.

Although Staddon is not clear on this point, it appears that some of these internal states can be further characterized as "mechanisms." A mechanism is defined as a "real-time rule of operation that describes the behavior" (p. 60). One example of a mechanism is a "satiation signal" that changes over time to control feeding, and another involves two transient memory systems that change over time and explain habituation. Neither internal states nor mechanisms are limited to the status of simple covert responses and stimuli.

DIFFERENCES WITH COGNITIVISM

Given this theoretical behaviorism with inferred theoretical states and mechanisms, how is Staddon's behaviorism different from contemporary cognitivism with its multitude of inferred theoretical processes? If Staddon is critical of radical behaviorism, he is equally so of cognitivism. He judges that cognitivism has been a scientific failure. According to him, cognitivism has ignored motivation in favor of representation and has reified its constructs.

In contrast to cognitivism, he denies the necessity of the distinction between performance and competence, he does not assume that inferred internal mechanisms are mental (i.e., what is reported by introspection), and his theory neither incorporates everyday mental language such as "expectation" nor seeks guidance from phenomenology or physiological plausibility. More basically, Staddon rejects as not useful the whole computation-representation model that underlies cognitivism, noting that to say that a system is "computing" or "representing" is merely one of many possible interpretations of a series of observed events.

Rather than limit theory to the tra-

ditional mental categories of the computation-representation model, Staddon advocates theorizing guided by only one basic principle—parsimony. His theoretical behaviorism makes use of any theoretical constructs that are both useful and parsimonious. As one example of such theorizing, he recommends neural network connectionism. In this model, behavior is the result of the combined activities of a large number of unintelligent nonsymbolic mechanisms, with no internal representation of the external world necessary. Thus, the model is S-R associationism with the addition of “nodes,” or states, that can be associated with one another, as well as with stimuli. That this model was adopted first by the field of artificial intelligence and only much later by behaviorism, its natural heir, Staddon attributes to Skinner’s “bizarre” opposition to theoretical entities. (For a more comprehensive and technical demonstration of Staddon’s theorizing, one should consult his forthcoming book, *Adaptive Dynamics: The Theoretical Analysis of Behavior*.)

CONSCIOUSNESS

As in any comprehensive system of psychology, Staddon’s new behaviorism must address the fundamental question of consciousness. Whereas Skinner’s behaviorism was “radical” in that consciousness itself was understood in terms of private stimuli and responses, Staddon reverts to an earlier behaviorist gambit and excludes phenomenological experience from the domain of his new behaviorism. Verbal reports of experience are included, but as behavior only.

This move considerably eases Staddon’s burden in explaining contemporary research that threatens to bring back consciousness into psychology, because he needs to explain only the verbal behavior rather than the qualities of the subjective experience. For example, Staddon interprets the phenomenon of color phi as the subject

giving equivalent responses to two sets of stimuli: The first is the set of external stimuli that give rise to color phi, and the second is the set of external stimuli typically described by the verbal report in color phi (i.e., an external single spot really moving and changing color at the midpoint). The quale of subjective experience is thus studied as the organism’s system of categorizing, measured by observable verbal responses.

Consciousness, then, for Staddon is a property of our system for self-report, not necessarily the initiator of action. This implies that some actions can be initiated by brain subsystems outside the “consciousness” subsystem and can thus occur “unconsciously.” This, in turn, enables Staddon to explain a number of paradoxical findings from experiments on brain stimulation that some commentators have seen as threatening to return dualism to psychology.

A NEW BEHAVIORISM?

Most of the empirical claims of Staddon’s proposed New Behaviorism already are the new behaviorism. Most contemporary behaviorists understand that contiguity and contingency are too simple to account for the selective effects of reinforcement, and they accept the importance of predictiveness. Similarly, the elicitive effects of reinforcement, the interaction of operant and respondent processes, instinctive drift, and the significance of what Staddon calls “variation” are all well documented and have necessitated changes in Skinner’s original conception of the operant. Although Staddon is also probably right that, contrary to Skinner, punishment is effective in changing behavior, it is not clear that Skinner has lost this battle. Applied behavior analysts still resist the use of punishment in clinical practice, preferring positive reinforcement. They do so not out of ignorance but perhaps out of an appreciation for the diffuse long-term negative effects of punishment, effects

that researchers underestimate because they are hard to measure experimentally.

These challenges to the basic science underlying radical behaviorism also call into question Skinner's many extrapolations from the laboratory to complex human affairs. Staddon is severely critical of Skinner's "premature extrapolations." But when is an extrapolation "premature"? That depends on one's purposes. If Skinner's extrapolations are taken as complete explanations of complex human activities, then of course they are harmful as well as premature because they create a false sense of understanding and are probably false in their details. However, if their purpose is to suggest hypotheses to guide future research, only time will tell if they are successful. Moreover, if their purpose is to persuade people of the power and promise of a particular conception of psychology, then they already have been successful, drawing many into the orbit of behaviorism. Staddon is unimpressed by Skinner's interpretations; others, including myself, have been inspired by them.

This issue of persuasion also arises with respect to Staddon's criticism of Skinner's opposition to the notions of freedom, responsibility, and autonomy. Yes, Staddon makes a philosophically sound argument that these notions are compatible with determinism. However, Skinner's purposes were not those of a philosopher debating at the faculty club. His goal was propaganda, to persuade society of the importance of behavioral control in social policy. Which is more effective at social persuasion—a subtle counterintuitive philosophical argument few will understand or a revolutionary call to abandon traditional values in order to preserve world peace, natural resources, and ultimately our culture? Although Staddon may be correct philosophically, it is not yet clear that Skinner made a tactical error, relative to his own purposes, in adopting his more radical stance.

Similarly, Staddon is undoubtedly

correct that the survival criterion is not a perfect guide for social policy. Without an advanced behavioral technology we are ignorant of how to control behavior fully, the results of that control, and how conditions will change in the future. But does this mean that we should not currently make use of our limited knowledge? In Woody Allen's "Sleeper," people in the distant future find that tobacco smoking is actually beneficial for health. Should we let this theoretical possibility inhibit our using social controls such as cigarette taxes and advertising to reduce smoking behavior? Staddon believes that examples like this are the minority, and he favors caution when using survival as a justification for major social changes. His cautionary message is sensible, but again, it is not obvious that he is right.

In the theoretical realm, Staddon is also a conservative in that much of what he advocates is a return to earlier forms of behaviorism. His preference for proximal-cause explanations is a call for a revised and more sophisticated classical behaviorism. Yet, in this call for a return, Staddon may be ignoring the success of "teleological" behaviorism. The matching law, maximization, and melioration are all examples of teleological principles that have been highly effective in organizing research programs, formulating theories, and explaining behavior in a parsimonious way. Staddon has not made a convincing case for the need to go beyond this approach.

Similarly, his advocacy for internal states, mechanisms, and connectionism is a return to S-R behaviorism and the theoretical constructs of the Hull-Spence tradition. However, Staddon may overestimate his differences with radical behaviorism on this issue. Despite Skinner's many explicit disavowals of theoretical inference, his system of behavior analysis did include intervening variables (e.g., reflex strength, response strength, emotion, deprivation, satiation) and hypothetical constructs (e.g., covert stimuli and responses).

Furthermore, because Staddon's internal states and mechanisms are all well-defined intervening variables, a determined instrumentalist could argue that they are all replaceable by terms in equations relating behavior to the environment. If so, the remaining difference on this issue between radical behaviorism and Staddon's theoretical behaviorism is whether these mathematical terms should be given names such as "satiation signal" and "transient memory systems." Certainly these names, pregnant with extramathematical meaning, suggest interesting hypotheses, both behavioral and physiological, and may be helpful in formulating research and theory. They also convey to us a deeper sense of understanding. In opposing such names, Skinner argued, however, that they tend to be reified and divert our attention away from behavior, the real subject matter. Thus, to some extent, this debate is over stylistic preference, and to some extent, over differing intuitions about how behavioral science progresses.

CONCLUSION

Because so much of this book is a sustained polemic against Skinner's radical behaviorism, it will be of greatest interest to those already familiar with Skinnerian ideas. Nevertheless, for those interested in a sound modern behaviorism, it can be found in this book, although not easily. One must extricate Staddon's views on social policy and psychological theory from his criticisms of radical behaviorism and cognitivism. The result is a version of behaviorism that can account for contemporary experimental findings and answer much of the criticism leveled at behaviorism. Characterizing this version, however, is difficult.

There is nothing in the social policy positions that dictates the theoretical positions and vice versa. To some extent, the Darwinian metaphor loosely integrates many of the ideas. To some extent, Staddon's conservatism underlies the program.

Staddon's personality threads through this book in other ways as well. His editor has thankfully given him wide latitude, and his dry wit (puns included) is frequently apparent. But the reader should be forewarned that Staddon's tone with regard to Skinner is unkind and unnecessarily harsh. His objections often descend to the personal. For example, in criticizing Skinner's position on punishment, Staddon says, "By abandoning his objections to punishment, Skinner would have destroyed the uniqueness of his position, which is something he was always at pains to maintain, most obviously through the adoption of an idiosyncratic vocabulary" (p. 111).

Despite the tone, the book is an outstanding example of modern behaviorist thought. Staddon has updated behaviorism with respect to contemporary experimental research, social policy, and philosophy of science. Anyone who attempts to evaluate behaviorism will have to come to terms with this comprehensive and intellectually powerful book. Staddon is familiar not only with behaviorist research but also with modern advances in cognitive science, philosophy, and social theory. He has combined his erudition and his penetrating analytic skills with his personal ideology to create an important work that is conservative—cautious about social change, suspicious of behavioral control, and opposed to grand revolutionary schemes, be they theoretical extrapolations or utopian social plans. Readers who share this disposition will find his arguments congenial; those who do not will not be persuaded.