Psychological Behaviorism: Origins of the Human Mind

Authored by

mohammad looti

June 16, 2026

Precursors: Wilhelm Wundt, Ivan Pavlov

Wundt is often called "the father of experimental psychology." He conceived the subject matter of psychology to be "experience in its relations to the subject" (Wundt 1897: 3). The science of experience he envisaged was supposed to be chemistry like: introspected experiential data were to be analyzed; the basic constituents of conscious experience thus identified; and the patterns and laws by which these basic constituents combine to constitute more complex conscious experiences (e.g., emotions) described. Data were to be acquired and analyzed by trained introspective Observers. While the analysis of experience was supposed to be a self-contained enterprise, Wundt -- originally trained as a physiologist -- fully expected that the structures and processes introspective analysis uncovered in experience would parallel structures and processes physiological investigation revealed in the central nervous system. Introspectionism, as the approach was called, soon spread, and laboratories sprang up in the United States and elsewhere, aiming "to investigate the facts of consciousness, its combinations and relations," so as to "ultimately discover the laws which govern these relations and combinations" (Wundt 1912: 1). The approach failed primarily due to the unreliability of introspective Observation. Introspective "experimental" results were not reliably reproducible by outside laboratories: Observers from different laboratories failed to agree, for instance, in their Observation (or failure to Observe) imageless thoughts (to cite one notorious controversy).

Pavlov's successful experimental discovery the laws of classical conditioning (as they came to be called), by way of contrast, provided positive inspiration for Watson's Behaviorist manifesto. Pavlov's stimulus-response model of explanation is also paradigmatic to much later behavioristic thought. In his famous experiments Pavlov paired presentations to dogs of an unconditioned stimulus (food) with an initially neutral stimulus (a ringing bell). After a number of such joint presentations, the unconditional responseto food (salivation) becomes conditioned to the bell: salivation occurs upon the ringing of the bell alone, in the absence of food. In accord with Pavlovian theory, then, given an animal's conditioning historybehavioral responses (e.g., salivation) can be predicted to occur or not, and be controlled (made to occur or not), on the basis of laws of conditioning, answering to the stimulus-response pattern:

S -> R

Everything adverted to here is publicly observable, even measurable; enabling Pavlov to experimentally investigate and formulate laws concerning temporal sequencing and delay effects, stimulus intensity effects, and stimulus generalization (opening doors to experimental investigation of animal perception and discrimination).

Edward Thorndike, in a similar methodological vein, proposed "that psychology may be, at least in part, as independent of introspection as physics" (Thorndike 1911: 5) and pursued experimental investigations of animal intelligence. In experimental investigations of puzzle-solving by cats and

other animals, he established that speed of solution increased gradually as a result of previous puzzle exposure. Such results, he maintained, support the hypothesis that learning is a result of habits formed through trial and error, and Thorndike formulated "laws of behavior," describing habit formation processes, based on these results. Most notable among Thorndike's laws (presaging Skinnerian operant conditioning) is his Law of Effect:

Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur; those which are accompanied or closely followed by discomfort to the animal will, other things being equal, have their connections with that situation weakened, so that, when it recurs, they will be less likely to occur. The greater the satisfaction or discomfort, the greater the strengthening or weakening of the bond . (Thorndike 1911)

In short, rewarded responses tend to be reinforced and punished responses eliminated. His methodological innovations (particularly his "puzzle-box") facilitated objective quantitative data collection and provided a paradigm for Behaviorist research methods to follow (especially the "Skinner box").

Early Behaviorism: John B. Watson

Watson coined the term "Behaviorism" as a name for his proposal to revolutionize the study of human psychology in order to put it on a firm experimental footing. In opposition to received philosophical opinion, to the dominant Introspectionist approach in psychology, and (many said) to common sense, Watson (1912) advocated a radically different approach. Where received "wisdom" took conscious experience to be the very stuff of minds and hence the (only) appropriate object of psychological investigation, Watson advocated an approach that led, scientifically, "to the ignoring of consciousness" and the illegitimacy of "making consciousness a special object of observation." He proposed, instead, that psychology should "take as a starting point, first the observable fact that organisms, man and animal alike, do adjust themselves to their environment" and "secondly, that certain stimuli lead the organisms to make responses." Whereas Introspectionism had, in Watson's estimation, miserably failed in its attempt to make experimental science out of subjective experience, the laboratories of animal psychologists, such as Pavlov and Thorndike, were already achieving reliably reproducible results and discovering general explanatory principles. Consequently, Watson -- trained as an "animal man" himself -- proposed, "making behavior, not consciousness, the objective point of our attack" as the key to putting the study of human psychology on a similar scientific footing. Key it proved to be. Watson's revolution was a smashing success. Introspectionism languished, behaviorism flourished, and considerable areas of our understanding of human psychology (particularly with regard to learning) came within the purview of experimental investigation along broadly behavioristic lines. Notably, also, Watson

foreshadows Skinner's ban on appeals to inner (central nervous) processes, seeming to share the Skinnerian sentiment "that because so little is known about the central nervous system, it serves as the last refuge of the soul in psychology" (Zuriff 1985: 80). Watson is, consequently, loath to hypothesize central processes, going so far as to speculate that thought occurs in the vocal tract, and is -- quite literally -- subaudible talking to oneself (Watson 1920).

Intermediaries: Edward Tolman and Clark Hull

Tolman and Hull were the two most noteworthy figures of the movement's middle years. Although both accepted the S-R framework as basic, Tolman and Hull were far more willing than Watson to hypothesize internal mechanisms or "intervening variables" mediating the S-R connection. In this regard their work may be considered precursory to cognitivism, and each touches on important philosophical issues besides. Tolman's purposive behaviorism attempts to explain goal-directed or purposive behavior, focusing on large, intact, meaningful behavior patterns or "molar" behavior (e.g., kicking a ball) as opposed to simple muscle movements or "molecular" behavior (e.g., various flexings of leg muscles); regarding the molecular level as too far removed from our perceptual capacities and explanatory purposes to provide suitable units for meaningful behavioral analysis. For Tolman, stimuli play a cognitive role as signals to the organism, leading to the formation of "cognitive maps" and to "latent learning" in the absence of reinforcement.

Overall, the stimuli which are allowed in are not connected by just simple one-to-one switches to the outgoing responses. Rather the incoming impulses are usually worked over and elaborated in the central control room into a tentative cognitive-like map of the environment. And it is this tentative map, indicating routes and paths and environmental relationships, which finally determines what responses, if any, the animal will finally make. (Tolman 1948: 192)

Clark Hull undertook the ambitious program of formulating an exhaustive theory of such mechanisms intervening between stimuli and responses: the theory was to take the form of a hypothetical-deductive system of basic laws or "postulates" enabling the prediction of behavioral responses (as "output variables") on the basis of external stimuli ("input variables") plus internal states of the organism ("intervening variables"). Including such organismic "intervening" variables (O) in the predictive/explanatory laws results in the following revised explanatory schema:

S & O -> R

The intervening O-variables Hull hypothesized included drive and habit strength. Attributes of, and relations among, these variables are what the postulates describe: further attributes and relationships were derived as theorems and corollaries from the basic postulates. Hull's student, Edward Spence, attempted to carry on with the program, without lasting success. Expected gains in predictive-explanatory scope and precision were not achieved and, with hindsight, it is easy to

see that such an elaborate theoretical superstructure, built on such slight observational-experimental foundations, was bound to fall. Hull's specific proposals are presently more historical curiosities than live hypotheses. Nevertheless, currently prevalent cognitivist approaches share Hull's general commitment to internal mechanisms.

B. F. Skinner: Radical Behaviorism

Skinner's self-described "radical behaviorist" approach is radical in its insistence on extending behaviorist strictures against inward experiential processes to include inner physiological ones as well. The scientific nub of the approach is a concept of operant conditioning indebted to Thorndike's "Law of Effect." Operants (e.g., bar-presses or key-pecks) are units of behavior an organism (e.g., a rat or pigeon) occasionally emits "spontaneously" prior to conditioning. In operant conditioning, operants followed byreinforcement (e.g., food or water) increase in frequency and come under control of discriminative stimuli (e.g., lights or tones) preceding the response. By increasingly judicious reinforcement of increasingly close approximations, complex behavioral sequences are shaped. On Skinner's view, high-level human behavior, such as speech, is the end result of such shaping. Prolonged absence of reinforcement leads to extinction of the response.

Many original and important Skinnerian findings -- e.g., that constantly reinforced responses extinguish more rapidly than intermittently reinforced responses -- concern the effects of differing schedules of reinforcement. Skinner notes the similarity of operant behavioral conditioning to natural evolutionary selection: in each case apparently forward-looking or goal-directed developments are explained (away) by a preceding course of environmental "selection" among randomly varying evolutionary traits or, in the psychological case, behavioral tricks. The purposiveness which Tolman's molar behavioral description assumes, radical behaviorism thus claims to explain. Likewise, Skinner questions the explanatory utility of would-be characterizations of inner processes (such as Hull's): such processes, being behavior themselves (though inner), are more in need of explanation themselves, Skinner holds, than they are fit to explain outward behavior. By "dismissing mental states and processes," Skinner maintains, radical behaviorism "directs attention to the ... history of the individual and to the current environment where the real causes of behavior are to be found" (Skinner 1987: 75). On this view, "if the proper attention is paid to the variables controlling behavior and an appropriate behavioral unit is chosen, orderliness appears directly in the behavior and the postulated theoretical processes become superfluous" (Zuriff: 88). Thus understood, Skinner's complaint about inner processes "is not that they do not exist, but that they are not relevant" (Skinner 1953) to the prediction, control, and experimental analysis of behavior.

Skinner stressed prediction and control as his chief explanatory desiderata, and on this score he boasts that "experimental analysis of behaviour" on radical behaviorist lines "has led to an effective technology, applicable to education, psychotherapy, and the design of cultural practices in general"

(Skinner 1987: 75). Even the most strident critics of radical behaviorism, I believe, must accord it some recognition in these connections. Behavior therapy (based on operant principles) has proven effective in treating phobias and addictions; operant shaping is widely and effectively used in animal training; and behaviorist instructional methods have proven effective -- though they may have become less fashionable -- in the field of education. Skinnerian Behaviorism can further boast of significantly advancing our understanding of stimulus generalization and other important learning-and-perception related phenomena and effects. Nevertheless, what was delivered was less than advertised. In particular, Skinner's attempt to extend the approach to the explanation of high-grade human behavior failed, making Noam Chomsky's dismissive (1959) review of Skinner's book, Verbal Behavior, something of a watershed. On Chomsky's diagnosis, not only had Skinner's attempt at explaining verbal behavior failed, it had to fail given the insufficiency of the explanatory devices Skinner allowed: linguistic competence (in general) and language acquisition (in particular), Chomsky argued, can only be explained as expressions of innate mechanisms -- presumably, computational mechanisms. For those in the "behavioral sciences" already chaffing under the severe methodological constraints Skinnerian orthodoxy imposed, the transition to "cognitive science" was swift and welcome. By 1985 Zuriff would write, "the received wisdom of today is that behaviorism has been refuted, its methods have failed, and it has little to offer modern psychology" (Zuriff 1985: 278). Subsequent developments, however, suggest that matters are not that simple.

Behaviorism

Loosely speaking, behaviorism is an attitude. Strictly speaking, behaviorism is a doctrine. A behaviorist, so understood, is a psychological theorist who demands behavioral evidence for any psychological hypothesis. For such a person, there is no knowable difference between two states of mind unless there is a demonstrable difference in the behavior associated with each state. Arguably, there is nothing truly exciting about behaviorism loosely understood. It enthrones behavioral evidence, an arguably inescapable practice in psychological science. Not so behaviorism the doctrine. This entry is about the doctrine, not the attitude. Behaviorism, the doctrine, has caused considerable excitation among both advocates and critics. Behaviorism, the doctrine, is committed in its fullest and most complete sense to the truth of the following three sets of claims.

Psychology is the science of behavior. Psychology is not the science of mind.
Behavior can be described and explained without making ultimate reference to mental events or to internal psychological processes. The sources of behavior are external (in the environment), not internal (in the mind, in the head).
In the course of theory development in psychology, if, somehow, mental terms or concepts are deployed in describing or explaining behavior, then either (a) these terms or concepts should be eliminated and replaced by behavioral terms or (b) they can and should be translated or

paraphrased into behavioral concepts.

The three sets of claims are logically distinct. Moreover, taken independently, each helps to form a type of behaviorism. "Methodological" behaviorism is committed to the truth of (1). "Psychological" behaviorism is committed to the truth of (2). "Analytical" behaviorism (also known as "philosophical" or "logical" behaviorism) is committed to the truth of the sub-statement in (3) that mental terms or concepts can and should be translated into behavioral concepts.

Other nomenclature is sometimes used to classify behaviorisms. Georges Rey (1997, p. 96), for example, classifies behaviorisms as methodological, analytical, and radical, where "radical" is Rey's term for what I am classifying as psychological behaviorism. I reserve the term "radical" for the psychological behaviorism of B. F. Skinner. Skinner employs the expression "radical behaviorism" to describe his brand of behaviorism or his philosophy of behaviorism (see Skinner 1974, p. 18). In the classification scheme used in this entry, radical behaviorism is a sub-type of psychological behaviorism, primarily, although it combines all three types of behaviorism (methodological, analytical, and psychological).

Three Types of Behaviorism

Methodological behaviorism is a normative theory about the scientific conduct of psychology. It claims that psychology should concern itself with the behavior of organisms (human and nonhuman animals). Psychology should not concern itself with mental states or events or with constructing internal information processing accounts of behavior. According to methodological behaviorism, reference to mental states, such as an animal's beliefs or desires, adds nothing to what psychology can and should understand about the sources of behavior. Mental states are private entities which, given the necessary publicity of science, do not form proper objects of empirical study. Methodological behaviorism is a dominant theme in the writings of John Watson (1878-1958).

Psychological behaviorism is a research program within psychology. It purports to explain human and animal behavior in terms of external physical stimuli, responses, learning histories, and (for certain types of behavior) reinforcements. Psychological behaviorism is present in the work of Ivan Pavlov (1849-1936), Edward Thorndike (1874-1949), as well as Watson. Its fullest and most influential expression is B. F. Skinner's work on schedules of reinforcement.

To illustrate, consider a food-deprived rat in an experimental chamber. If a particular movement, such as pressing a lever when a light is on, is followed by the presentation of food, then the likelihood of the rat's pressing the lever when hungry, again, and the light is on, is increased. Such presentations are reinforcements, such lights are (discriminative) stimuli, such lever pressings are responses, and such trials or associations are learning histories.

Analytical or logical behaviorism is a theory within philosophy about the meaning or semantics of mental terms or concepts. It says that the very idea of a mental state or condition is the idea of a behavioral disposition or family of behavioral tendencies, evident in how a person behaves in one situation rather than another. When we attribute a belief, for example, to someone, we are not saying that he or she is in a particular internal state or condition. Instead, we are characterizing the person in terms of what he or she might do in particular situations or environmental interactions. Analytical behaviorism may be found in the work of Gilbert Ryle (1900-76) and the later work of Ludwig Wittgenstein (1889-51) (if perhaps not without controversy in interpretation, in Wittgenstein's case). More recently, the philosopher-psychologist U. T. Place (1924-2000) advocated a brand of analytical behaviorism restricted to intentional or representational states of mind, such as beliefs, which Place took to constitute a type, although not the only type, of mentality (see Graham and Valentine 2004). Arguably, a version of analytical or logical behaviorism may also be found in the work of Daniel Dennett on the ascription of states of consciousness via a method he calls 'heterophenomenology' (Dennett 2005, pp. 25-56). (See also Melser 2004.)

Roots of Behaviorism

Each of methodological, psychological, and analytical behaviorism has historical foundations. Analytical behaviorism traces its historical roots to the philosophical movement known as Logical Positivism (see Smith 1986). Logical positivism proposes that the meaning of statements used in science be understood in terms of experimental conditions or observations that verify their truth. This positivist doctrine is known as "verificationism." In psychology, verificationism underpins or grounds analytical behaviorism, namely, the claim that mental concepts refer to behavioral tendencies and so must be translated into behavioral terms.

Analytical behaviorism helps to avoid substance dualism. Substance dualism is the doctrine that mental states take place in a special, non-physical mental substance (the immaterial mind). By contrast, for analytical behaviorism, the belief that I have as I arrive on time for a 2pm dental appointment, namely, that I have a 2pm appointment, is not the property of a mental substance. Believing is a family of tendencies of my body. In addition, for an analytical behaviorist, we cannot identify the belief about my arrival independently of that arrival or other members of this family of tendencies. So, we also cannot treat it as the cause of the arrival. Cause and effect are, as Hume taught, conceptually distinct existences. Believing that I have a 2pm appointment is not distinct from my arrival and so cannot be part of the causal foundations of arrival.

Psychological behaviorism's historical roots consist, in part, in the classical associationism of the British Empiricists, foremost John Locke (1632-1704) and David Hume (1711-76). According to classical associationism, intelligent behavior is the product of associative learning. As a result of associations or pairings between perceptual experiences or stimulations on the one hand, and ideas or thoughts on the other, persons and animals acquire knowledge of their environment and

how to act. Associations enable creatures to discover the causal structure of the world. Association is most helpfully viewed as the acquisition of knowledge about relations between events. Intelligence in behavior is a mark of such knowledge.

Classical associationism relied on introspectible entities, such as perceptual experiences or stimulations as the first links in associations, and thoughts or ideas as the second links. Psychological behaviorism, motivated by experimental interests, claims that to understand the origins of behavior, reference to stimulations (experiences) should be replaced by reference to stimuli (physical events in the environment), and that reference to thoughts or ideas should be eliminated or displaced in favor of reference to responses (overt behavior, motor movement). Psychological behaviorism is associationism without appeal to mental events.

Don't human beings talk of introspectible entities, thoughts, feelings, and so on, even if these are not recognized by behaviorism or best understood as behavioral tendencies? Psychological behaviorists regard the practice of talking about one's own states of mind, and of introspectively reporting those states, as potentially useful data in psychological experiments, but as not presupposing the metaphysical subjectivity or non-physical presence of those states. There are different sorts of causes behind introspective reports, and psychological behaviorists take these and other elements of introspection to be amenable to behavioral analysis. (For additional discussion, see Section 5 of this entry). (See, for comparison, Dennett's method of heterophenomenology; Dennett 1991, pp. 72-81)

The task of psychological behaviorism is to specify types of association, understand how environmental events control behavior, discover and elucidate causal regularities or laws or functional relations which govern the formation of associations, and predict how behavior will change as the environment changes. The word "conditioning" is commonly used to specify the process involved in acquiring new associations. Animals in so-called "operant" conditioning experiments are not learning to, for example, press levers. Instead, they are learning about the relationship between events in their environment, for example, that a particular behavior, pressing the lever, causes food to appear.

In its historical foundations, methodological behaviorism shares with analytical behaviorism the influence of positivism. One of the main goals of positivism was to unify psychology with natural science. Watson wrote that "psychology as a behaviorist views it is a purely objective experimental branch of natural science. Its theoretical goal is ... prediction and control" (1913, p. 158). Watson also wrote of the purpose of psychology as follows: "To predict, given the stimulus, what reaction will take place; or, given the reaction, state what the situation or stimulus is that has caused the reaction" (1930, p. 11).

Though logically distinct, methodological, psychological, and analytical behaviorisms often are found in one behaviorism. Skinner's radical behaviorism combines all three forms of behaviorism. It

follows analytical strictures (at least loosely) in paraphrasing mental terms behaviorally, when or if they cannot be eliminated from explanatory discourse. In Verbal Behavior (1957) and elsewhere, Skinner tries to show how mental terms can be given behavioral interpretations. In About Behaviorism (1974) he says that when mental terminology cannot be eliminated it can be "translated into behavior" (p. 18, Skinner brackets the expression with his own double quotes).
Radical behaviorism is concerned with the behavior of organisms, not with internal processing. So, it is a form of methodological behaviorism. Finally, radical behaviorism understands behavior as a reflection of frequency effects among stimuli, which means that it is a form of psychological behaviorism.

Popularity of Behaviorism

Behaviorism of one sort or another was an immensely popular research program or methodological commitment among students of behavior from about the second decade of the twentieth century through its middle decade, at least until the beginnings of the cognitive science revolution (see Bechtel, Abrahamsen, and Graham, 1998, pp. 15-17). In addition to Ryle and Wittgenstein, philosophers with sympathies for behaviorism included Carnap (1932-33), Hempel (1949), and Quine (1960). Quine, for example, took a behaviorist approach to the study of language. Quine claimed that the notion of psychological or mental activity has no place in a scientific account of either the origins or the meaning of speech. To talk in a scientifically disciplined manner about the meaning of an utterance is to talk about stimuli for the utterance, its so-called "stimulus meaning". Hempel (1949) claimed that "all psychological statements that are meaningful ... are translatable into statements that do not involve psychological concepts," but only concepts for physical behavior (p. 18).

Among psychologists behaviorism was even more popular than among philosophers. In addition to Pavlov, Skinner, Thorndike, and Watson, the list of behaviorists among psychologists included, among others, E. C. Tolman (1886-1959), C. L. Hull (1884-52), and E. R. Guthrie (1886-1959). Tolman, for example, wrote that "everything important in psychology ... can be investigated in essence through the continued experimental and theoretical analysis of the determiners of rat behavior at a choice point in a maze" (1938, p. 34).

Behaviorists created journals, organized societies, and founded psychology graduate programs reflective of behaviorism. Behaviorists organized themselves into different types of research clusters, whose differences stemmed from such factors as varying approaches to conditioning and experimentation. Some clusters were named as follows: "the experimental analysis of behavior", "behavior analysis", "functional analysis", and, of course, "radical behaviorism". These labels sometimes were responsible for the titles of behaviorism's leading societies and journals, including the Society for the Advancement of Behavior Analysis (SABA), and the Journal of the Experimental Analysis of Behavior (begun in 1958) as well as the Journal of Applied Behavior Analysis (begun in

1968).

Behaviorism generated a type of therapy, known as behavior therapy (see Rimm and Masters 1974; Erwin 1978). It developed behavior management techniques for autistic children (see Lovaas and Newsom 1976) and token economies for the management of chronic schizophrenics (see Stahl and Leitenberg 1976). It fueled discussions of how best to understand the behavior of nonhuman animals, the relevance of laboratory study to the natural environmental occurrence of behavior, and whether there is built-in associative bias in learning (see Schwartz and Lacey 1982).

Behaviorism stumbled upon various critical difficulties with some of its commitments. One difficulty is confusion about the effects of reinforcement on behavior (see Gallistel 1990). In its original sense, a stimulus such as food is a reinforcer only if its presentation increases the frequency of a response in a type of associative conditioning known as operant conditioning. A problem with this definition is that it defines reinforcers as stimuli that change behavior. The presentation of food, however, may have no observable effect on response frequency even in cases in which an animal is food deprived. Rather, response frequency can be associated with an animal's ability to identify and remember temporal or spatial properties of the circumstances in which a stimulus is presented. This and other difficulties prompted changes in behaviorism's commitments and new directions of research. One recent and fresh direction has been the study of the role of short term memory in contributing to reinforcement effects on the so-called trajectory of behavior (see Killeen 1994).

Another stumbling block, in the case of analytical behaviorism, is the fact that the behavioral sentences that are intended to offer the behavioral paraphrases of mental terms almost always use mental terms themselves (see Chisholm 1957). In the example of my belief that I have a 2pm dental appointment, one must also speak of my desire to arrive at 2pm, otherwise the behavior of arriving at 2pm could not count as believing that I have a 2pm appointment. The term "desire" is a mental term. Critics have charged that we can never escape from using mental terms in the characterization of the meaning of mental terms. This suggests that mental discourse cannot be displaced by behavioral discourse. At least it cannot be displaced term-by-term. Perhaps analytical behaviorists need to paraphrase a whole swarm of mental terms at once so as to recognize the presumption that the attribution of any one such mental term presupposes the application of others (see Rey 1997, p. 154-5).

Behaviorism: Positive Aspects

Why would anyone be a behaviorist? There are three main reasons (see also Zuriff 1985).

The first is epistemic. Warrant or evidence for saying, at least in the third person case, that an animal or person is in a certain mental state, for example, possesses a certain belief, is grounded

in behavior, understood as observable behavior. Moreover, the conceptual space between the claim that behavior warrants the attribution of belief and the claim that believing consists in behavior is a short and in some ways appealing step. If we look, for example, at how people are taught to use mental concepts and terms--terms like "believe", "desire", and so on--conditions of use appear inseparably connected with behavioral tendencies in certain circumstances. If mental state attribution bears a special connection with behavior, it is tempting to say that mentality just consists in behavioral tendencies.

The second reason can be expressed as follows: One major difference between mentalistic (mental states in-the-head) and associationist or conditioning accounts of behavior is that mentalistic accounts tend to have a strong nativist bent. This is true even though there may be nothing inherently nativist about mentalistic accounts (see Cowie 1998).

Mentalistic accounts tend to assume, and sometimes even explicitly to embrace (see Fodor 1981), the hypothesis that the mind possesses at birth or innately a set of procedures or internally represented processing rules which are deployed when learning or acquiring new responses. Behaviorism, by contrast, is anti-nativist. Behaviorism, therefore, appeals to theorists who deny that there are innate rules by which organisms learn. To Skinner and Watson organisms learn without being innately or pre-experientially provided with implicit procedures by which to learn. Learning does not consist, at least initially, in rule-governed behavior. Learning is what organisms do in response to stimuli. For a behaviorist an organism learns, as it were, from its successes and mistakes. "Rules," says Skinner (1984a), "are derived from contingencies, which specify discriminative stimuli, responses, and consequences" (p. 583). (See also Dennett 1978).

Much contemporary work in cognitive science on the set of models known as connectionist or parallel distributed processing (PDP) models seems to share behaviorism's anti-nativism about learning. PDP takes an approach to learning which is response oriented rather than rule-governed and this is because, like behaviorism, it has roots in associationism (see Bechtel 1985; compare Graham 1991 with Maloney 1991). Whether PDP models ultimately are or must be anti-nativist depends upon what counts as native or innate rules (Bechtel and Abrahamsen 1991, pp. 103-105).

The third reason for behaviorism's appeal, popular at least historically, is related to its disdain for reference to inner mental or information processing as explanatory causes of behavior. The disdain is most vigorously exemplified in the work of Skinner. Skinner's skepticism about explanatory references to mental innerness may be described as follows.

Suppose we try to explain the public behavior of a person by describing how they represent or think about their situation. Suppose we mean by this that a person classifies and analyzes the environment or situation that they are in. They conceive or think of it a certain way, not as bare, as items without attributes, but as things, as trees, as people, as walruses, walls, and wallets. Suppose, we say, a person never merely interacts with their environment; but rather interacts with

their environment as they perceive, see, or represent it. So, for example, thinking of something as a wallet, a person reaches for it. Perceiving something as a walrus, they back away from it. Classifying it as a wall, they approach it. So understood, behavior is endogenously produced movement, behavior that has its causal origin within the person who thinks of their situation a certain way.

Skinner would object to such claims. He would object not because he believes that the eye is innocent or that inner or endogenous activity does not occur. He would object because he believes that behavior must be explained in terms that do not themselves presuppose the very thing that is explained. This is behavior. The outside (public) behavior of a person is not accounted for by referring to the inside (inner processing, cognitive activity) behavior of the person (say, his or her classifying or analyzing their environment) if, therein, the behavior of the person is unexplained. "The objection," wrote Skinner, "to inner states is not that they do not exist, but that they are not relevant in a functional analysis" (Skinner 1953, p. 35). 'Not relevant' means, for Skinner, explanatorily circular or regressive.

Skinner charges that since mental activity is a form of behavior (albeit inner), the only non-regressive, non-circular way to explain behavior is to appeal to something non-behavioral. This non-behavioral something is environmental stimuli and an organism's interactions with, and reinforcement from, the environment.

So, the third reason for behaviorism's appeal is that it tries to avoid circular, regressive explanations of behavior. It aims to refrain from accounting for one type of behavior (overt) in terms of another type of behavior (covert), all the while, in some sense, leaving behavior unexplained.

It should be noted that Skinner's views about explanation and the purported circularity of explanation by reference to inner processing are both extreme and scientifically contestable, and that many who have self-identified as behaviorists including Guthrie, Tolman, and Hull, or continue to work within the tradition, broadly understood, including Killeen (1987) and Rescorla (1990), take exception to much that Skinner has said about explanatory references to innerness. It should also be noted that Skinner's derisive attitude towards explanatory references to mental innerness stems, in part, not just from fears of explanatory regression but from his conviction that if the language of psychology is permitted to refer to internal processing, this goes some way towards permitting talk of immaterial mental substances, agents endowed with contra-causal free will, and little persons (homunculi) within bodies. Each of these Skinner takes to be incompatible with a scientific worldview (see Skinner 1971; see also Day 1976). Finally, it must be noted that Skinner's aversion to explanatory references to innerness is not an aversion to inner mental states or processes per se. He readily admits that they exist. Skinner countenances talk of inner events provided that they are treated in the same manner as public or overt responses. An adequate science of behavior, he claims, must describe events taking place within the skin of the organism

as part of behavior itself (see Skinner 1976). "So far as I am concerned," he wrote in 1984 in a special issue of Behavioral and Brain Sciences devoted to his work, "whatever happens when we inspect a public stimulus is in every respect similar to what happens when we introspect a private one" (Skinner 1984b, p. 575; compare Graham 1984, pp. 558-9).

Skinner does not have much to say about just how inner or covert behavior (like thinking, classifying, and analyzing) can be described in the same manner as public or overt behavior. But his idea seems roughly to be this. Just as we may describe overt behavior or motor movement in terms of concepts like stimulus, response, conditioning, reinforcement, and so on, so we may deploy the very same terms in describing inner or covert behavior. One thought may reinforce another thought. An act of analysis may serve as a stimulus for an effort at classification. And so on. Purely 'mentalistic' activities may be at least roughly parsed in terms of behavioral concepts -- a topic to be revisited later in the entry (in the 7th Section).

Skinner's Social Worldview

Skinner is the only major figure in the history of behaviorism to offer a socio-political world view based on his commitment to behaviorism. Skinner constructed a theory as well as narrative picture in Walden Two (1948) of what an ideal human society would be like if designed according to behaviorist principles (see also Skinner 1971). Skinner's social worldview illustrates both his aversion to free will, to homunculi, to dualism as well as his reasons for claiming that a person's history of environmental interactions controls his or her behavior.
One remarkable feature of human behavior which Skinner deliberately rejects is that people creatively make their own environments (see Chomsky 1971, Black 1973). The world is as it is, in part, because we make it that way. Skinner protests that "it is in the nature of an experimental analysis of human behavior that it should strip away the functions previously assigned to autonomous man and transfer them one by one to the controlling environment" (1971, p. 198).

Critics have raised several objections to the Skinnerian social picture. One of the most persuasive, and certainly one of the most frequent, adverts to Skinner's vision of the ideal human society. It is a question asked of the fictional founder of Walden Two, Frazier, by the philosopher Castle. It is the question of what is the best social mode of existence for a human being. Frazier's, and therein Skinner's, response to this question is both too general and incomplete. Frazier/Skinner speaks of the values of health, friendship, relaxation, rest, and so forth. However, these values are hardly the detailed basis of a social system.

There is a notorious difficulty in social theory of specifying the appropriate level of detail at which a blueprint for a new and ideal society must be presented (see Arnold 1990, pp. 4-10). Skinner identifies the behavioristic principles and learning incentives that he hopes will reduce systematic injustices in social systems. He also describes a few practices (concerning child rearing and the

like) that are intended to contribute to human happiness. However, he offers only the haziest descriptions of the daily lives of Walden Two citizens and no suggestions for how best to resolve disputes about alternative ways of life that are prima facie consistent with behaviorist principles (see Kane 1996, p. 203). He gives little or no serious attention to the crucial general problem of inter-personal conflict resolution and to the role of institutional arrangements in resolving conflicts.

In an essay which appeared in The Behavior Analyst (1985), nearly forty years after the publication of Walden Two, Skinner, in the guise of Frazier, tried to clarify his characterization of ideal human circumstances. He wrote that in the ideal human society "people just naturally do the things they need to do to maintain themselves ... and treat each other well, and they just naturally do a hundred other things they enjoy doing because they do not have to do them" (p. 9). However, of course, doing a hundred things humans enjoy doing means only that Walden Two is vaguely defined, not that its culturally instituted habits and the character of its institutions merit emulation.
The incompleteness of Skinner's description of the ideal human society or life is so widely acknowledged that one might wonder if actual experiments in Walden Two living could lend useful detail to his blueprint. More than one such social experiment has been conducted. Perhaps the most interesting (in part because the community has evolved away from its Skinnerian influences) is the Twin Oaks Community in Virginia in the U.S.A., which can be indirectly explored via the Internet (see Other Internet Resources).

Criticism

Behaviorism has lost strength and influence. It is dismissed by cognitive scientists developing intricate internal information processing models of cognition. Its laboratory routines are neglected by cognitive ethologists and ecological psychologists convinced that its methods are irrelevant to studying how animals and persons behave in their natural and social environment. Its traditional relative indifference towards neuroscience and deference to environmental contingencies is rejected by neuroscientists sure that direct study of the brain is the only way to understand the causes of behavior.

But by no means has behaviorism disappeared. Robust elements of behaviorism survive in both behavior therapy and laboratory-based animal learning theory (of which more below). In the metaphysics of mind, too, behavioristic themes survive in the approach to mind known as functionalism. Functionalism defines states of mind as states that play particular causal-functional roles in animals or systems in which they occur. Paul Churchland writes of functionalism as follows: "The essential or defining feature of any type of mental states is the set of causal relations it bears to ... bodily behavior" (1984, p. 36). This functionalist notion is similar to the behaviorist idea that reference to behavior and to stimulus/response relations enters centrally and essentially into any account of what it means for a creature to behave or to be subject, in the scheme of analytical or logical behaviorism, to the attribution of mental states. Elements, however, are

elements. Behaviorism is no longer a dominating research program.

Why has the influence of behaviorism declined?

The deepest and most complex reason for behaviorism's decline in influence is its commitment to the thesis that behavior can be explained without reference to non-behavioral mental (cognitive, representational, or interpretative) activity. Behavior can be explained just by reference to its "functional" (Skinner's term) relation to or co-variation with the environment and to the animal's history of environmental interaction. Neurophysiological and neurobiological conditions, for Skinner, sustain or implement these functional relations. But they do not serve as ultimate or independent sources of behavior. Behavior, Skinner (1953) wrote, cannot be accounted for "while staying wholly inside ; eventually we must turn to forces operating upon the organism from without." "Unless there is a weak spot in our causal chain so that the second link is not lawfully determined by the first , or the third by the second, the first and third links must be lawfully related." (p. 35) "Valid information about the second link may throw light on this relationship but can in no way alter it." (ibid.) It is "external variables of which behavior is a function." (ibid.)

Skinner was no triumphalist about neuroscience. Neuroscience, for him, more or less just identifies organismic physical processes that underlie animal/environment interactions. Therein, it rides evidential or epistemic piggyback on radical behaviorism's prior description of those interactions. "The organism", he says, "is not empty, and it cannot adequately be treated simply as a black box" (1976, p. 233). "Something is done today which affects the behavior of the organism tomorrow" (p. 233). Neuroscience describes inside-the-box mechanisms that permit today's reinforcing stimulus to affect tomorrow's behavior. The neural box is not empty, but it is unable, except in cases of malfunction or breakdown, to disengage the animal from past patterns of behavior that have been reinforced. It cannot exercise independent or non-environmentally countervailing authority over behavior.

For many critics of behaviorism it seems obvious that, at a minimum, the occurrence and character of behavior (especially human behavior) does not depend primarily upon an individual's reinforcement history, although that is a factor, but on the fact that the environment or learning history is represented by an individual and how (the manner in which) it is represented. The fact that the environment is represented by me constrains or informs the functional relations that hold between my behavior and the environment and may, from an anti-behaviorist perspective, partially disengage my behavior from its conditioning or reinforcement history. No matter, for example, how tirelessly and repeatedly I have been reinforced for pointing to or eating ice cream, such a history is impotent if I just don't see a potential stimulus as ice cream or represent it to myself as ice cream or if I desire to hide the fact that something is ice cream from others. My conditioning history, narrowly understood as unrepresented by me, is behaviorally less important than the environment or my learning history as represented or interpreted by me.

Similarly, for many critics of behaviorism, if representationality comes between environment and behavior, this implies that Skinner is too restrictive or limited in his attitude towards the role of brain mechanisms in producing or controlling behavior. The brain is no mere passive memory bank of behavior/environment interactions (see Roediger and Goff 1998). The central nervous system, which otherwise sustains my reinforcement history, contains systems or neurocomputational sub-systems that implement or encode whatever representational content the environment has for me. It is also an active interpretation machine or semantic engine, often critically performing environmentally untethered and behavior controlling tasks. Such talk of representation or interpretation, however, is a perspective from which behaviorism--most certainly in Skinner--wished to depart.

One defining feature of traditional behaviorism is that it tried to free psychology from having to theorize about how animals and persons represent their environment. This was important, historically, because it seemed that behavior/environment connections are a lot clearer and more manageable experimentally than internal representations. Unfortunately, for behaviorism, it's hard to imagine a more restrictive rule for psychology than one which prohibits hypotheses about representational storage and processing. Stephen Stich, for example, complains against Skinner that "we now have an enormous collection of experimental data which, it would seem, simply cannot be made sense of unless we postulate something like" information processing mechanisms in the heads of organisms (1998, p. 649).

A second reason for rejecting behaviorism is that some features of mentality--some elements in the inner processing of persons--have characteristic 'qualia' or presentationally immediate or phenomenal qualities. To be in pain, for example, is not merely to produce appropriate pain behavior under the right environmental circumstances, but it is to experience a 'like-thisness' to the pain (as something dull or sharp, perhaps). A purely behaviorist creature, a 'zombie', as it were, may engage in pain behavior, including beneath the skin pain responses, yet completely lack whatever is qualitatively distinctive of and proper to pain (its painfulness). (See also Graham 1998, pp. 47-51 and Graham and Horgan 2000. On the scope of the phenomenal in human mentality, see Graham, Horgan, and Tienson 2009).

The philosopher-psychologist U.T. Place, although otherwise sympathetic to the application of behaviorist ideas to matters of mind, argued that phenomenal qualia cannot be analyzed in behaviorist terms. He claimed that qualia are neither behavior nor dispositions to behave. "They make themselves felt," he said, "from the very moment that the experience of whose qualia they are" comes into existence (2000, p. 191; reprinted in Graham and Valentine 2004). They are instantaneous features of processes or events rather than dispositions manifested over time. Qualitative mental events (such as sensations, perceptual experiences, and so on), for Place, undergird dispositions to behave rather than count as dispositions. Indeed, it is tempting to postulate that the qualitative aspects of mentality affect non-qualitative elements of internal

processing, and that they, for example, contribute to arousal, attention, and receptivity to associative conditioning.

The third reason for rejecting behaviorism is connected with Noam Chomsky. Chomsky has been one of behaviorism's most successful and damaging critics. In a review of Skinner's book on verbal behavior (see above), Chomsky (1959) charged that behaviorist models of language learning cannot explain various facts about language acquisition, such as the rapid acquisition of language by young children, which is sometimes referred to as the phenomenon of "lexical explosion." A child's linguistic abilities appear to be radically underdetermined by the evidence of verbal behavior offered to the child in the short period in which he or she expresses those abilities. By the age of four or five (normal) children have an almost limitless capacity to understand and produce sentences which they have never heard before. Chomsky also argued that it seems just not to be true that language learning depends on the application of reinforcement. A child does not, as an English speaker in the presence of a house, utter "house" repeatedly in the presence of reinforcing elders. Language as such seems to be learned without, in a sense, being explicitly taught or taught in detail, and behaviorism doesn't offer an account of how this could be so. Chomsky's own speculations about the psychological realities underlying language development included the hypothesis that the rules or principles underlying linguistic behavior are abstract (applying to all human languages) and innate (part of our native psychological endowment as human beings). When put to the test of uttering a grammatical sentence, a person, for Chomsky, has a virtually infinite number of possible responses available, and the only way in which to understand this virtually infinite generative capacity is to suppose that a person possesses a powerful and abstract innate grammar (underlying whatever competence he or she may have in one or more particular natural languages).

The problem to which Chomsky refers, which is the problem of behavioral competence and thus performance outstripping individual learning histories, goes beyond merely the issue of linguistic behavior in young children. It appears to be a fundamental fact about human beings that our behavior and behavioral capacities often surpass the limitations of individual reinforcement histories. Our history of reinforcement is often too impoverished to determine uniquely what we do or how we do it. Much learning, therefore, seems to require pre-existing or innate representational structures or principled constraints within which learning occurs. (See also Brewer 1974, but compare with Bates et al. 1998 and Cowie 1998).

Is the case against behaviorism definitive? Decisive? Paul Meehl noted more than three decades ago that theories in psychology seem to disappear not under the force of decisive refutation but rather because researchers lose interest in their theoretical orientations (Meehl 1978). One implication of Meehl's thesis is that a once popular "Ism", not having been decisively refuted, may restore some of its former prominence if it mutates or transforms itself so as to incorporate responses to criticisms. What may this mean for behaviorism? It may mean that some version of

the doctrine might rebound.

Skinner claimed that neural activities subserve behavior/environment relations and that the organism's contribution to these relations does not reduce to neurophysiological properties. But this does not mean that behaviorism cannot gain useful alliance with neuroscience. Reference to brain structures (neurobiology, neurochemistry, and so on) may help in explaining behavior even if such reference does not ultimately displace reference to environmental contingencies in a behaviorist account.

Such is a lesson of animal modeling in which behaviorist themes still enjoy currency. Animal models of addiction, habit and instrumental learning are particularly noteworthy because they bring behavioral research into closer contact than did traditional psychological behaviorism with research on the brain mechanisms underlying reinforcement, especially positive reinforcement (West 2006, pp. 91-108). One result of this contact is the discovery that sensitized neural systems responsible for heightened reinforcement value or strength can be dissociated from the hedonic utility or pleasurable quality of reinforcement (see Robinson and Berridge 2003). The power of a stimulus to reinforce behavior may be independent of whether it is a source of pleasure. Focus on brain mechanisms underlying reinforcement also forms the centerpiece of one of the most active research programs in current neuroscience, so-called neuroeconomics, which weds study of the brain's reward systems with models of valuation and economic decision making (see Montague and Berns 2002; Nestler and Malenka 2004; Ross et al 2008). Behaviorism may do well to purchase some of neuroeconomic's conceptual currency, especially since some advocates of the program see themselves as behaviorists in spirit if not stereotypical letter and honor the work of a number of theorists in the behavioristic tradition of the experimental analysis of behavior, such as George Ainslie, Richard Herrnstein and Howard Rachlin, on how patterns of behavior relate to patterns of reward or reinforcement (see Ross et al. 2008, especially p. 10). One assumption in neuroeconomics is that full explanations of organism/environmental interactions will combine facts about such things as reinforcement schedules with appeal to neurocomputational modeling and to the neurochemistry and neurobiology of reinforcement.

Other potential sources of renewal? The continued popularity of behavior therapy is noteworthy because it offers a potential domain of application for the regimen of logical or analytical behaviorism. Early versions of behavior therapy sought to apply results from Skinnerian or Pavlovian conditioning paradigms to human behavior problems. No minds should be spoken of; just behavior--stimuli, responses, and reinforcement. Therapy shapes behavior not thought. Successive generations of behavior therapy have relaxed those conceptual restrictions. Advocates refer to themselves as cognitive behavior therapists (e.g. Mahoney, 1974; Meichenbaum, 1977). Clients' behavior problems are described by referring to their beliefs, desires, intentions, memories, and so on. Even the language of self-reflexive thought and belief (so-called 'meta-cognition') figures in some accounts of behavioral difficulties and interventions (Wells 2000). One goal of such

language is to encourage clients to monitor and self-reinforce their own behavior. Self-reinforcement is an essential feature of behavioral self-control (Rachlin 2000; Ainslie 2001).

It may be wondered whether cognitive behavior therapy is consistent with behaviorist doctrine. Much depends on how beliefs and desires are understood. If beliefs and desires are understood as states that somehow spill out into the environment and are individuated in terms of their non-mentalistic role in organism/environment interactions, this would be consistent with traditional behaviorist doctrine. It would reflect the principle of logical or analytical behaviorism that if mental terms are to be used in the description and explanation of behavior, they must be defined or paraphrased in non-mental behavioral terms. Prospects for belief/desire individuation in non-mental terms may look doubtful (see Horgan and Graham forthcoming). But the topic is open for further exploration.

Conclusion

In 1977 Willard Day, a behavioral psychologist and founding editor of the journal Behaviorism (which now is known as Behavior and Philosophy), published Skinner's Why I am not a cognitive psychologist" (Skinner 1977). Skinner began the paper by stating that "the variables of which human behavior is a function lie in the environment" (p. 1). Skinner ended by remarking that "cognitive constructs give ... a misleading account of what" is inside a human being (p. 10)

More than a decade earlier, in 1966 Carl Hempel had announced his defection from behaviorism:

In order to characterize ... behavioral patterns, propensities, or capacities ... we need not only a suitable behavioristic vocabulary, but psychological terms as well. (p. 110)

Hempel had come to believe that it is a mistake to imagine that human behavior can be understood exclusively in non-mental, behavioristic terms. Contemporary psychology and philosophy largely share Hempel's conviction that the explanation of behavior cannot omit invoking a creature's representation of its world. Psychology must use psychological terms. Behavior without cognition is blind. Psychological theorizing without reference to internal cognitive processing is explanatorily impaired. To say this, of course, is not to a priori preclude that behaviorism will recover some of its prominence. Just how to conceive of cognitive processing (even where to locate it) remains a heated subject of debate (see Melser 2004; see also Levy 2007, pp. 29-64). But if behaviorism is to recover some of its prominence, this recovery may require a reformulation of its doctrines that is attune to developments (like that of neuroeconomics) in neuroscience as well as in novel therapeutic orientations.

Skinner's vantage point on behaviorism mates the science of behavior with the language of organism/environment interactions. But we don't just run and mate and walk and eat. We think, classify, analyze, and theorize. In addition to our outer behavior, we have highly complex inner

lives, wherein we are active, imaginatively, in our heads, all the while often remaining as stuck as posts, as still as stones. To figure out how all that maps into the Country of Behaviorism remains the "ism's" still incompletely charted territory.