REASON AT WORK

Introductory Readings in Philosophy

STEVEN M. CAHN
PATRICIA KITCHER
GEORGE SHER

56
Behaviorism

B. F. SKINNER

For biographical information about B. F. Skinner, see reading 41.

In these excerpts from Science and Human Behavior, *Skinner presents a comprehensive account of the theory of behaviorism and argues for the superiority of his approach to various other ways of understanding "mental phenomena." Skinner claims that it is both unscientific and fruitless to try to explain human behavior by reference to inner causes—thoughts or neural activity, for example. Instead we should seek the causes of human behavior outside the individual in the environment. Skinner maintains that all human behavior can be explained by reference to three different stimulus-response relations. A narrow range of human behaviors, such as tearing in the presence of onions, can be described as "unconditioned reflexes." These activities are simply automatic, untrained responses to the presence of certain stimuli. A somewhat wider class of behaviors is accurately described as "conditioned reflexes." The best known ixample is the salivation evoked in Pavlov's dogs by the ringing of a bell, after the bell-ringing had been paired with the natural or unconditioned stimulus of food. According to Skinner, the vast majority of human behavior is the result of "operant conditioning." Operant conditioning takes place when a particular behavior is rewarded or "reinforced." So, for example, Skinner would account for the fact that you are studying philosophy not by reference to your desires or goals, but by looking back in your life history for occasions on which your family or your culture rewarded you for this type of activity.*

THE TERMS "cause" and "effect" are no longer widely used in science. They have been associated with so many theories of the structure and operation of the universe that they mean more than scientists want to say. The terms which replace them, however, refer to the same factual core. A "cause" becomes a "change in an independent variable" and an "effect" a "change in a dependent variable." The old "cause-and-effect connection" becomes a "functional relation." The new terms do not suggest *how* a cause causes its effect: they merely assert that different events tend to occur together in a certain order. This is important, but it is not crucial. There is no particular danger in using "cause" and "effect" in an informal discussion if we are always ready to substitute their more exact counterparts.

We are concerned, then, with the causes of human behavior. We want to know why men behave as they do. Any condition or event which can be shown to have an effect upon behavior must be taken into account. By discovering

and analyzing these causes we can predict behavior; to the extent that we can manipulate them, we can control behavior.

There is a curious inconsistency in the zeal with which the doctrine of personal freedom has been defended, because men have always been fascinated by the search for causes. The spontaneity of human behavior is apparently no more challenging than its "why and wherefore." So strong is the urge to explain behavior that men have been led to anticipate legitimate scientific inquiry and to construct highly implausible theories of causation. This practice is not unusual in the history of science. The study of any subject begins in the realm of superstition. The fanciful explanation precedes the valid. Astronomy began as astrology; chemistry as alchemy. The field of behavior has had, and still has, its astrologers and alchemists. A long history of prescientific explanation furnishes us with a fantastic array of causes which have no function other than to supply spurious answers to questions which must otherwise go unanswered in the early stages of a science.

. . .

Inner "Causes"

Every science has at some time or other looked for causes of action inside the things it has studied. Sometimes the practice has proved useful, sometimes it has not. There is nothing wrong with an inner explanation as such, but events which are located inside a system are likely to be difficult to observe. For this reason we are encouraged to assign properties to them without justification. Worse still, we can invent causes of this sort without fear of contradiction. The motion of a rolling stone was once attributed to its *vis viva*. The chemical properties of bodies were thought to be derived from the *principles* or *essences* of which they were composed. Combustion was explained by the *phlogiston* inside the combustible object. Wounds healed and bodies grew well because of a *vis medicatrix*. It has been especially tempting to attribute the behavior of a living organism to the behavior of an inner agent, as the following examples may suggest.

Neural Causes. The layman uses the nervous system as a ready explanation of behavior. The English language contains hundreds of expressions which imply such a causal relationship. At the end of a long trial we read that the jury shows signs of *brain fag,* that the *nerves* of the accused are *on edge,* that the wife of the accused is on the verge of a *nervous breakdown,* and that his lawyer is generally thought to have lacked the *brains* needed to stand up to the prosecution. Obviously, no direct observations have been made of the nervous systems of any of these people. Their "brains" and "nerves" have been invented on the spur of the moment to lend substance to what might otherwise seem a superficial account of their behavior.

. . .

Eventually a science of the nervous system based upon direct observation rather than inference will describe the neural states and events which immediately precede instances of behavior. We shall know the precise neurological

conditions which immediately precede, say, the response, "No, thank you." These events in turn will be found to be preceded by other neurological events, and these in turn by others. This series will lead us back to events outside the nervous system and, eventually, outside the organism. In the chapters which follow we shall consider external events of this sort in some detail. We shall then be better able to evaluate the place of neurological explanations of behavior. However, we may note here that we do not have and may never have this sort of neurological information at the moment it is needed in order to predict a specific instance of behavior. It is even more unlikely that we shall be able to alter the nervous system directly in order to set up the antecedent conditions of a particular instance. The causes to be sought in the nervous system are, therefore, of limited usefulness in the prediction and control of specific behavior.

Psychic inner causes. An even more common practice is to explain behavior in terms of an inner agent which lacks physical dimensions and is called "mental" or "psychic." The purest form of the psychic explanation is seen in the animism of primitive peoples. From the immobility of the body after death it is inferred that a spirit responsible for movement has departed. The *enthusiastic* person is, as the etymology of the word implies, energized by a "god within." It is only a modest refinement to attribute every feature of the behavior of the physical organism to a corresponding feature of the "mind" or of some inner "personality." The inner man is regarded as driving the body very much as the man at the steering wheel drives a car. The inner man wills an action, the outer executes it. The inner loses his appetite, the outer stops eating. The inner man wants and the outer gets. The inner has the impulse which the outer obeys.

It is not the layman alone who resorts to these practices, for many reputable psychologists use a similar dualistic system of explanation. The inner man is sometimes personified clearly, as when delinquent behavior is attributed to a "disordered personality," or he may be dealt with in fragments, as when behavior is attributed to mental processes, faculties, and traits. Since the inner man does not occupy space, he may be multiplied at will. It has been argued that a single physical organism is controlled by several psychic agents and that its behavior is the resultant of their several wills. The Freudian concepts of the ego, superego, and id are often used in this way. They are frequently regarded as nonsubstantial creatures, often in violent conflict, whose defeats or victories lead to the adjusted or maladjusted behavior of the physical organism in which they reside.

Direct observation of the mind comparable with the observation of the nervous system has not proved feasible. It is true that many people believe that they observe their "mental states" just as the physiologist observes neural events, but another interpretation of what they observe is possible. . . . Introspective psychology no longer pretends to supply direct information about events which are the causal antecedents, rather than the mere accompaniments, of behavior. It defines its "subjective" events in ways which strip them of any usefulness in a causal analysis. The events appealed to in early mentalistic explanations of behavior have remained beyond the reach of observation. Freud insisted upon

this by emphasizing the role of the unconscious—a frank recognition that important mental processes are not directly observable. The Freudian literature supplies many examples of behavior from which unconscious wishes, impulses, instincts, and emotions are inferred. Unconscious thought-processes have also been used to explain intellectual achievements. Though the mathematician may feel that he knows "how he thinks," he is often unable to give a coherent account of the mental processes leading to the solution of a specific problem. But any mental event which is unconscious is necessarily inferential, and the explanation is therefore not based upon independent observations of a valid cause.

The fictional nature of this form of inner cause is shown by the ease with which the mental process is discovered to have just the properties needed to account for the behavior. When a professor turns up in the wrong classroom or gives the wrong lecture, it is because his *mind* is, at least for the moment, *absent*. If he forgets to give a reading assignment, it is because it has slipped his *mind* (a hint from the class may *remind* him of it). He begins to tell an old joke but pauses for a moment, and it is evident to everyone that he is trying to make up his *mind* whether or not he has already used the joke that term. His lectures grow more tedious with the years, and questions from the class confuse him more and more, because his *mind* is failing. What he says is often disorganized because his *ideas* are confused. He is occasionally unnecessarily emphatic because of the force of his *ideas*. When he repeats himself, it is because he has an *idée fixe;* and when he repeats what others have said, it is because he borrows his *ideas*. Upon occasion there is nothing in what he says because he lacks *ideas*. In all this it is obvious that the mind and the ideas, together with their special characteristics, are being invented on the spot to provide spurious explanations. A science of behavior can hope to gain very little from so cavalier a practice. Since mental or psychic events are asserted to lack the dimensions of physical science, we have an additional reason for rejecting them.

Conceptual inner causes. The commonest inner causes have no specific dimensions at all, either neurological or psychic. When we say that a man eats *because* he is hungry, smokes a great deal *because* he has the tobacco habit, fights *because* of the instinct of pugnacity, behaves brilliantly *because* of his intelligence, or plays the piano well *because* of his musical ability, we seem to be referring to causes. But on analysis these phrases prove to be merely redundant descriptions. A single set of facts is described by the two statements: "He eats" and "He is hungry." A single set of facts is described by the two statements: "He smokes a great deal" and "He has the smoking habit." A single set of facts is described by the two statements: "He plays well" and "He has musical ability." The practice of explaining one statement in terms of the other is dangerous because it suggests that we have found the cause and therefore need search no further. Moreover, such terms as "hunger," "habit," and "intelligence" convert what are essentially the properties of a process or relation into what appear to be things. Thus we are unprepared for the properties eventually to be discovered in the behavior itself and continue to look for something which may not exist.

The Variables of Which Behavior is a Function

The practice of looking inside the organism for an explanation of behavior has tended to obscure the variables which are immediately available for a scientific analysis. These variables lie outside the organism, in its immediate environment and in its environmental history. They have a physical status to which the usual techniques of science are adapted, and they make it possible to explain behavior as other subjects are explained in science. These independent variables are of many sorts and their relations to behavior are often subtle and complex, but we cannot hope to give an adequate account of behavior without analyzing them.

Consider the act of drinking a glass of water. This is not likely to be an important bit of behavior in anyone's life, but it supplies a convenient example. We may describe the topography of the behavior in such a way that a given instance may be identified quite accurately by any qualified observer. Suppose now we bring someone into a room and place a glass of water before him. Will he drink? There appear to be only two possibilities: either he will or he will not. But we speak of the *chances* that he will drink, and this notion may be refined for scientific use. What we want to evaluate is the *probability* that he will drink. This may range from virtual certainty that drinking will occur to virtual certainty that it will not. The very considerable problem of how to measure such a probability will be discussed later. For the moment, we are interested in how the probability may be increased or decreased.

Everyday experience suggests several possibilities, and laboratory and clinical observations have added others. It is decidedly not true that a horse may be led to water but cannot be made to drink. By arranging a history of severe deprivation we could be "absolutely sure" that drinking would occur. In the same way we may be sure that the glass of water in our experiment will be drunk. Although we are not likely to arrange them experimentally, deprivations of the necessary magnitude sometimes occur outside the laboratory. We may obtain an effect similar to that of deprivation by speeding up the excretion of water. For example, we may induce sweating by raising the temperature of the room or by forcing heavy exercise, or we may increase the excretion of urine by mixing salt or urea in food taken prior to the experiment. It is also well known that loss of blood, as on a battlefield, sharply increases the probability of drinking. On the other hand, we may set the probability at virtually zero by inducing or forcing our subject to drink a large quantity of water before the experiment.

If we are to predict whether or not our subject will drink, we must know as much as possible about these variables. If we are to induce him to drink, we must be able to manipulate them. In both cases, moreover, either for accurate prediction or control, we must investigate the effect of each variable quantitatively with the methods and techniques of a laboratory science.

Other variables may, of course, affect the result. Our subject may be "afraid" that something has been added to the water as a practical joke or for experi-

mental purposes. He may even "suspect" that the water has been poisoned. He may have grown up in a culture in which water is drunk only when no one is watching. He may refuse to drink simply to prove that we cannot predict or control his behavior. These possibilities do not disprove the relations between drinking and the variables listed in the preceding paragraphs; they simply remind us that other variables may have to be taken into account. We must know the history of our subject with respect to the behavior of drinking water, and if we cannot eliminate social factors from the situation, then we must know the history of his personal relations to people resembling the experimenter. Adequate prediction in any science requires information about all relevant variables, and the control of a subject matter for practical purposes makes the same demands.

Other types of "explanation" do not permit us to dispense with these requirements or to fulfill them in any easier way. It is of no help to be told that our subject will drink provided he was born under a particular sign of the zodiac which shows a preoccupation with water or provided he is the lean and thirsty type or was, in short, "born thirsty." Explanations in terms of inner states or agents, however, may require some further comment. To what extent is it helpful to be told, "He drinks because he is thirsty"? If to be thirsty means nothing more than to have a tendency to drink, this is mere redundancy. If it means that he drinks because of a state of thirst, an inner causal event is invoked. If this state is purely inferential—if no dimensions are assigned to it which would make direct observation possible—it cannot serve as an explanation. But if it has physiological or psychic properties, what role can it play in a science of behavior?

The physiologist may point out that several ways of raising the probability of drinking have a common effect: they increase the concentration of solutions in the body. Through some mechanism not yet well understood, this may bring about a corresponding change in the nervous system which in turn makes drinking more probable. In the same way, it may be argued that all these operations make the organism "feel thirsty" or "want a drink" and that such a psychic state also acts upon the nervous system in some unexplained way to induce drinking. In each case we have a causal chain consisting of three links: (1) an operation performed upon the organism from without—for example, water deprivation; (2) an inner condition—for example, physiological or psychic thirst; and (3) a kind of behavior—for example, drinking. Independent information about the second link would obviously permit us to predict the third without recourse to the first. It would be a preferred type of variable because it would be non-historic; the first link may lie in the past history of the organism, but the second is a current condition. Direct information about the second link is, however, seldom, if ever, available. Sometimes we infer the second link from the third: an animal is judged to be thirsty if it drinks. In that case, the explanation is spurious. Sometimes we infer the second link from the first: an animal is said to be thirsty if it has not drunk for a long time. In that case, we obviously cannot dispense with the prior history.

The second link is useless in the *control* of behavior unless we can manipulate it. At the moment, we have no way of directly altering neural processes at appropriate moments in the life of a behaving organism, nor has any way been discovered to alter a psychic process. We usually set up the second link through the first: we make an animal thirsty, in either the physiological or the psychic sense, by depriving it of water, feeding it salt, and so on. In that case, the second link obviously does not permit us to dispense with the first. Even if some new technical discovery were to enable us to set up or change the second link directly, we should still have to deal with those enormous areas in which human behavior is controlled through manipulation of the first link. A technique of operating upon the second link would increase our control of behavior, but the techniques which have already been developed would still remain to be analyzed.

The most objectionable practice is to follow the causal sequence back only as far as a hypothetical second link. This is a serious handicap both in a theoretical science and in the practical control of behavior. It is no help to be told that to get an organism to drink we are simply to "make it thirsty" unless we are also told how this is to be done. When we have obtained the necessary prescription for thirst, the whole proposal is more complex than it need be. Similarly, when an example of maladjusted behavior is explained by saying that the individual is "suffering from anxiety," we have still to be told the cause of the anxiety. But the external conditions which are then invoked could have been directly related to the maladjusted behavior. Again, when we are told that a man stole a loaf of bread because "he was hungry," we have still to learn of the external conditions responsible for the "hunger." These conditions would have sufficed to explain the theft.

The objection to inner states is not that they do not exist, but that they are not relevant in a functional analysis. We cannot account for the behavior of any system while staying wholly inside it; 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, then the first and third links must be lawfully related. If we must always go back beyond the second link for prediction and control, we may avoid many tiresome and exhausting digressions by examining the third link as a function of the first. Valid information about the second link may throw light upon this relationship but can in no way alter it.

A Functional Analysis

The external variables of which behavior is a function provide for what may be called a causal or functional analysis. We undertake to predict and control the behavior of the individual organism. This is our "dependent variable"—the effect for which we are to find the cause. Our "independent variables"—the causes of behavior—are the external conditions of which behavior is a function. Relations between the two—the "cause-and-effect relationships" in

behavior—are the laws of a science. A synthesis of these laws expressed in quantitative terms yields a comprehensive picture of the organism as a behaving system.

This must be done within the bounds of a natural science. We cannot assume that behavior has any peculiar properties which require unique methods or special kinds of knowledge. It is often argued that an act is not so important as the "intent" which lies behind it, or that it can be described only in terms of what it "means" to the behaving individual or to others whom it may affect. If statements of this sort are useful for scientific purposes, they must be based upon observable events, and we may confine ourselves to such events exclusively in a functional analysis. We shall see later that although such terms as "meaning" and "intent" appear to refer to properties of behavior, they usually conceal references to independent variables. This is also true of "aggressive," "friendly," "disorganized," "intelligent," and other terms which appear to describe properties of behavior but in reality refer to its controlling relations.

The independent variables must also be described in physical terms. An effort is often made to avoid the labor of analyzing a physical situation by guessing what it "means" to an organism or by distinguishing between the physical world and a psychological world of "experience." This practice also reflects a confusion between dependent and independent variables. The events affecting an organism must be capable of description in the language of physical science. It is sometimes argued that certain "social forces" or the "influences" of culture or tradition are exceptions. But we cannot appeal to entities of this sort without explaining how they can affect both the scientist and the individual under observation. The physical events which must then be appealed to in such an explanation will supply us with alternative material suitable for a physical analysis.

. . .

Reflex Action

Descartes had taken an important step in suggesting that some of the spontaneity of living creatures was only apparent and that behavior could sometimes be traced to action from without. The first clear-cut evidence that he had correctly surmised the possibility of external control came two centuries later in the discovery that the tail of a salamander would move when part of it was touched or pierced, even though the tail had been severed from the body. Facts of this sort are now familiar, and we have long since adapted our beliefs to take them into account. At the time the discovery was made, however, it created great excitement. It was felt to be a serious threat to prevailing theories of the inner agents responsible for behavior. If the movement of the amputated tail could be controlled by external forces, was its behavior when attached to the salamander of a different nature? If not, what about the inner causes which had hitherto been used to account for it? It was seriously suggested as an answer that the "will" must be coexistent with the body and that some part of it must invest any amputated part. But the fact remained that an external event had

been identified which could be substituted, as in Descartes's daring hypothesis, for the inner explanation.

The external agent came to be called a *stimulus*. The behavior controlled by it came to be called a *response*. Together they comprised what was called a *reflex*—on the theory that the disturbance caused by the stimulus passed to the central nervous system and was "reflected" back to the muscles. It was soon found that similar external causes could be demonstrated in the behavior of larger portions of the organism—for example, in the body of a frog, cat, or dog in which the spinal cord had been severed at the neck. Reflexes including parts of the brain were soon added, and it is now common knowledge that in the intact organism many kinds of stimulation lead to almost inevitable reactions of the same reflex nature. Many characteristics of the relation have been studied quantitatively. The time which elapses between stimulus and response (the "latency") has been measured precisely. The magnitude of the response has been studied as a function of the intensity of the stimulus. Other conditions of the organism have been found to be important in completing the account—for example, a reflex may be "fatigued" by repeated rapid elicitation.

The reflex was at first closely identified with hypothetical neural events in the so-called "reflex arc." A surgical division of the organism was a necessary entering wedge, for it provided a simple and dramatic method of analyzing behavior. But surgical analysis became unnecessary as soon as the principle of the stimulus was understood and as soon as techniques were discovered for handling complex arrangements of variables in other ways. By eliminating some conditions, holding others constant, and varying others in an orderly manner, basic lawful relations could be established without dissection and could be expressed without neurological theories.

The extension of the principle of the reflex to include behavior involving more and more of the organism was made only in the face of vigorous opposition. The reflex nature of the spinal animal was challenged by proponents of a "spinal will." The evidence they offered in support of a residual inner cause consisted of behavior which apparently could not be explained wholly in terms of stimuli. When higher parts of the nervous system were added, and when the principle was eventually extended to the intact organism, the same pattern of resistance was followed. But arguments for spontaneity, and for the explanatory entities which spontaneity seems to demand, are of such form that they must retreat before the accumulating facts. Spontaneity is negative evidence; it points to the weakness of a current scientific explanation, but does not in itself prove an alternative version. By its very nature, spontaneity must yield ground as a scientific analysis is able to advance. As more and more of the behavior of the organism has come to be explained in terms of stimuli, the territory held by inner explanations has been reduced. The "will" has retreated up the spinal cord, through the lower and then the higher parts of the brain, and finally, with the conditioned reflex, has escaped through the front of the head. At each stage, some part of the control of the organism has passed from a hypothetical inner entity to the external environment.

The Range of Reflex Action

A certain part of behavior, then, is elicited by stimuli, and our prediction of that behavior is especially precise. When we flash a light in the eye of a normal subject, the pupil contracts. When he sips lemon juice, saliva is secreted. When we raise the temperature of the room to a certain point, the small blood vessels in his skin enlarge, blood is brought nearer to the skin, and he "turns red." We use these relations for many practical purposes. When it is necessary to induce vomiting, we employ a suitable stimulus—an irritating fluid or a finger in the throat. The actress who must cry real tears resorts to onion juice on her handkerchief.

As these examples suggest, many reflex responses are executed by the "smooth muscles" (for example, the muscles in the walls of the blood vessels) and the glands. These structures are particularly concerned with the internal economy of the organism. They are most likely to be of interest in a science of behavior in the emotional reflexes to be discussed in Chapter X. Other reflexes use the "striped muscles" which move the skeletal frame of the organism. The "knee jerk" and other reflexes which the physician uses for diagnostic purposes are examples. We maintain our posture, either when standing still or moving about, with the aid of a complex network of such reflexes.

In spite of the importance suggested by these examples, it is still true that if we were to assemble all the behavior which falls into the pattern of the simple reflex, we should have only a very small fraction of the total behavior of the organism. This is not what early investigators in the field expected. We now see that the principle of the reflex was overworked. The exhilarating discovery of the stimulus led to exaggerated claims. It is neither plausible nor expedient to conceive of the organism as a complicated jack-in-the-box with a long list of tricks, each of which may be evoked by pressing the proper button. The greater part of the behavior of the intact organism is not under this primitive sort of stimulus control. The environment affects the organism in many ways which are not conveniently classed as "stimuli," and even in the field of stimulation only a small part of the forces acting upon the organism elicit responses in the invariable manner of reflex action. To ignore the principle of the reflex entirely, however, would be equally unwarranted.

Conditioned Reflexes

. . .

The difference between an unskilled conjecture and a scientific fact is not simply a difference in evidence. It had long been known that a child might cry before it was hurt or that a fox might salivate upon seeing a bunch of grapes. What Pavlov added can be understood most clearly by considering his history. Originally he was interested in the process of digestion, and he studied the conditions under which digestive juices were secreted. Various chemical substances in the mouth or in the stomach resulted in the reflex action of the

digestive glands. Pavlov's work was sufficiently outstanding to receive the Nobel Prize, but it was by no means complete. He was handicapped by a certain unexplained secretion. Although food in the mouth might elicit a flow of saliva, saliva often flowed abundantly when the mouth was empty. We should not be surprised to learn that this was called "psychic secretion." It was explained in terms which "any child could understand." Perhaps the dog was "thinking about food." Perhaps the sight of the experimenter preparing for the next experiment "reminded" the dog of the food it had received in earlier experiments. But these explanations did nothing to bring the unpredictable salivation within the compass of a rigorous account of digestion.

Pavlov's first step was to control conditions so that "psychic secretion" largely disappeared. He designed a room in which contact between dog and experimenter was reduced to a minimum. The room was made as free as possible from incidental stimuli. The dog could not hear the sound of footsteps in neighboring rooms or smell accidental odors in the ventilating system. Pavlov then built up a "psychic secretion" step by step. In place of the complicated stimulus of an experimenter preparing a syringe or filling a dish with food, he introduced controllable stimuli which could be easily described in physical terms. In place of the accidental occasions upon which stimulation might precede or accompany food, Pavlov arranged precise schedules in which controllable stimuli and food were presented in certain orders. Without influencing the dog in any other way, he could sound a tone and insert food into the dog's mouth. In this way he was able to show that the tone *acquired* its ability to elicit secretion, and he was also able to follow the process through which this came about. Once in possession of these facts, he could then give a satisfactory account of all secretion. He had replaced the "psyche" of psychic secretion with certain objective facts in the recent history of the organism.

The process of conditioning, as Pavlov reported it in his book *Conditioned Reflexes,* is a process of *stimulus substitution.* A previously neutral stimulus acquires the power to elicit a response which was originally elicited by another stimulus. The change occurs when the neutral stimulus is followed or "reinforced" by the effective stimulus. Pavlov studied the effect of the interval of time elapsing between stimulus and reinforcement. He investigated the extent to which various properties of stimuli could acquire control. He also studied the converse process, in which the conditioned stimulus loses its power to evoke the response when it is no longer reinforced—a process which he called "extinction."

The quantitative properties which he discovered are by no means "known to every child." And they are important. The most efficient use of conditioned reflexes in the practical control of behavior often requires quantitative information. A satisfactory theory makes the same demands. In dispossessing explanatory fictions, for example, we cannot be sure that an event of the sort implied by "psychic secretion" is not occasionally responsible until we can predict the exact amount of secretion at any given time. Only a quantitative description will make sure that there is no additional mental process in which the dog "associates the sound of the tone with the idea of food" or in which

it salivates because it "expects" food to appear. Pavlov could dispense with concepts of this sort only when he could give a complete quantitative account of salivation in terms of the stimulus, the response, and the history of conditioning.

Pavlov, as a physiologist, was interested in how the stimulus was converted into neural processes and in how other processes carried the effect through the nervous system to the muscles and glands. The subtitle of his book is *An Investigation of the Physiological Activity of the Cerebral Cortex*. The "physiological activity" was inferential. We may suppose, however, that comparable processes will eventually be described in terms appropriate to neural events. Such a description will fill in the temporal and spatial gaps between an earlier history of conditioning and its current result. The additional account will be important in the integration of scientific knowledge but will not make the relation between stimulus and response any more lawful or any more useful in prediction and control. Pavlov's achievement was the discovery, not of neural processes, but of important quantitative relations which permit us, regardless of neurological hypotheses, to give a direct account of behavior in the field of the conditioned reflex.

. . .

The Range of Conditioned Reflexes

Although the process of conditioning greatly extends the scope of the eliciting stimulus, it does not bring all the behavior of the organism within such stimulus control. According to the formula of stimulus substitution we must elicit a response before we can condition it. All conditioned reflexes are, therefore, based upon unconditioned reflexes. But we have seen that reflex responses are only a small part of the total behavior of the organism. Conditioning adds new controlling stimuli, but not new responses. In using the principle, therefore, we are not subscribing to a "conditioned-reflex theory" of all behavior.

. . .

Learning Curves

One of the first serious attempts to study the changes brought about by the consequences of behavior was made by E. L. Thorndike in 1898. His experiments arose from a controversy which was then of considerable interest. Darwin, in insisting upon the continuity of species, had questioned the belief that man was unique among the animals in his ability to think. Anecdotes in which lower animals seemed to show the "power of reasoning" were published in great numbers. But when terms which had formerly been applied only to human behavior were thus extended, certain questions arose concerning their meaning. Did the observed facts point to mental processes, or could these apparent evidences of thinking be explained in other ways? Eventually it became clear that the assumption of inner thought-processes was not required. Many years were to pass before the same question was seriously raised concerning human behavior, but Thorndike's experiments and his alternative explanation of reasoning in animals were important steps in that direction.

If a cat is placed in a box from which it can escape only by unlatching a door, it will exhibit many different kinds of behavior, some of which may be effective in opening the door. Thorndike found that when a cat was put into such a box again and again, the behavior which led to escape tended to occur sooner and sooner until eventually escape was as simple and quick as possible. The cat had solved its problem as well as if it were a "reasoning" human being, though perhaps not so speedily. Yet Thorndike observed no "thought-process" and argued that none was needed by way of explanation. He could describe his results simply by saying that a part of the cat's behavior was "stamped in" because it was followed by the opening of the door.

The fact that behavior is stamped in when followed by certain consequences, Thorndike called "The Law of Effect." What he had observed was that certain behavior occurred more and more readily in comparison with other behavior characteristic of the same situation. By noting the successive delays in getting out of the box and plotting them on a graph, he constructed a "learning curve." This early attempt to show a quantitative process in behavior, similar to the processes of physics and biology, was heralded as an important advance. It revealed a process which took place over a considerable period of time and which was not obvious to casual inspection. Thorndike, in short, had made a discovery. Many similar curves have since been recorded and have become the substance of chapters on learning in psychology texts.

Learning curves do not, however, describe the basic process of stamping in. Thorndike's measure—the time taken to escape—involved the elimination of other behavior, and his curve depended upon the number of different things a cat might do in a particular box. It also depended upon the behavior which the experimenter or the apparatus happened to select as "successful" and upon whether this was common or rare in comparison with other behavior evoked in the box. A learning curve obtained in this way might be said to reflect the properties of the latch box rather than of the behavior of the cat. The same is true of many other devices developed for the study of learning. The various mazes through which white rats and other animals learn to run, the "choice boxes" in which animals learn to discriminate between properties or patterns of stimuli, the apparatuses which present sequences of material to be learned in the study of human memory—each of these yields its own type of learning curve.

By averaging many individual cases, we may make these curves as smooth as we like. Moreover, curves obtained under many different circumstances may agree in showing certain general properties. For example, when measured in this way, learning is generally "negatively accelerated"—improvement in performance occurs more and more slowly as the condition is approached in which further improvement is impossible. But it does not follow that negative acceleration is characteristic of the basic process. Suppose, by analogy, we fill a glass jar with gravel which has been so well mixed that pieces of any given size are evenly distributed. We then agitate the jar gently and watch the pieces rearrange themselves. The larger move toward the top, the smaller toward the bottom. This process, too, is negatively accelerated. At first the mixture separates rap-

idly, but as separation proceeds, the condition in which there will be no further change is approached more and more slowly. Such a curve may be quite smooth and reproducible, but this fact alone is not of any great significance. The curve is the result of certain fundamental processes involving the contact of spheres of different sizes, the resolution of the forces resulting from agitation, and so on, but it is by no means the most direct record of these processes.

Learning curves show how the various kinds of behavior evoked in complex situations are sorted out, emphasized, and reordered. The basic process of the stamping in of a single act brings this change about, but it is not reported directly by the change itself.

Operant Conditioning

To get at the core of Thorndike's Law of Effect, we need to clarify the notion of "probability of response." This is an extremely important concept; unfortunately, it is also a difficult one. In discussing human behavior, we often refer to "tendencies" or "predispositions" to behave in particular ways. Almost every theory of behavior uses some such term as "excitatory potential," "habit strength," or "determining tendency." But how do we observe a tendency? And how can we measure one?

If a given sample of behavior existed in only two states, in one of which it always occurred and in the other never, we should be almost helpless in following a program of functional analysis. An all-or-none subject matter lends itself only to primitive forms of description. It is a great advantage to suppose instead that the *probability* that a response will occur ranges continuously between these all-or-none extremes. We can then deal with variables which, unlike the eliciting stimulus, do not "cause a given bit of behavior to occur" but simply make the occurrence more probable. We may then proceed to deal, for example, with the combined effect of more than one such variable.

The everyday expressions which carry the notion of probability, tendency, or predisposition describe the frequencies with which bits of behavior occur. We never observe a probability as such. We say that someone is "enthusiastic" about bridge when we observe that he plays bridge often and talks about it often. To be "greatly interested" in music is to play, listen to, and talk about music a good deal. The "inveterate" gambler is one who gambles frequently. The camera "fan" is to be found taking pictures, developing them, and looking at pictures made by himself and others. The "highly sexed" person frequently engages in sexual behavior. The "dipsomaniac" drinks frequently.

In characterizing a man's behavior in terms of frequency, we assume certain standard conditions: he must be able to execute and repeat a given act, and other behavior must not interfere appreciably. We cannot be sure of the extent of a man's interest in music, for example, if he is necessarily busy with other things. When we come to refine the notion of probability of response for scientific use, we find that here, too, our data are frequencies and that the conditions under which they are observed must be specified. The main technical problem in designing a controlled experiment is to provide for the observation

and interpretation of frequencies. We eliminate, or at least hold constant, any condition which encourages behavior which competes with the behavior we are to study. An organism is placed in a quiet box where its behavior may be observed through a one-way screen or recorded mechanically. This is by no means an environmental vacuum, but the organism will react to the features of the box in many ways; but its behavior will eventually reach a fairly stable level, against which the frequency of a selected response may be investigated.

To study the process which Thorndike called stamping in, we must have a "consequence." Giving food to a hungry organism will do. We can feed our subject conveniently with a small food tray which is operated electrically. When the tray is first opened, the organism will probably react to it in ways which interfere with the process we plan to observe. Eventually, after being fed from the tray repeatedly, it eats readily, and we are then ready to make this consequence contingent upon behavior and to observe the result.

We select a relatively simple bit of behavior which may be freely and rapidly repeated, and which is easily observed and recorded. If our experimental subject is a pigeon, for example, the behavior of raising the head above a given height is convenient. This may be observed by sighting across the pigeon's head at a scale pinned on the far wall of the box. We first study the height at which the head is normally held and select some line on the scale which is reached only infrequently. Keeping our eye on the scale we then begin to open the food tray very quickly whenever the head rises above the line. If the experiment is conducted according to specifications, the result is invariable: we observe an immediate change in the frequency with which the head crosses the line. We also observe, and this is of some importance theoretically, that higher lines are now being crossed. We may advance almost immediately to a higher line in determining when food is to be presented. In a minute or two, the bird's posture has changed so that the top of the head seldom falls below the line which we first chose.

When we demonstrate the process of stamping in in this relatively simple way, we see that certain common interpretations of Thorndike's experiment are superfluous. The expression "trial-and-error learning," which is frequently associated with the Law of Effect, is clearly out of place here. We are reading something into our observations when we call any upward movement of the head a "trial," and there is no reason to call any movement which does not achieve a specified consequence an "error." Even the term "learning" is misleading. The statement that the bird "learns that it will get food by stretching its neck" is an inaccurate report of what has happened. To say that it has acquired the "habit" of stretching its neck is merely to resort to an explanatory fiction, since our only evidence of the habit is the acquired tendency to perform the act. The barest possible statement of the process is this: we make a given consequence contingent upon certain physical properties of behavior (the upward movement of the head), and the behavior is then observed to increase in frequency.

It is customary to refer to any movement of the organism as a "response." The word is borrowed from the field of reflex action and implies an act which, so to speak, answers a prior event—the stimulus. But we may make an event

contingent upon behavior without identifying, or being able to identify, a prior stimulus. We did not alter the environment of the pigeon to *elicit* the upward movement of the head. It is probably impossible to show that any single stimulus invariably precedes this movement. Behavior of this sort may come under the control of stimuli, but the relation is not that of elicitation. The term "response" is therefore not wholly appropriate but is so well established that we shall use it in the following discussion.

A response which has already occurred cannot, of course, be predicted or controlled. We can only predict that *similar* responses will occur in the future. The unit of a predictive science is, therefore, not a response but a class of responses. The word "operant" will be used to describe this class. The term emphasizes the fact that the behavior *operates* upon the environment to generate consequences. The consequences define the properties with respect to which responses are called similar. The term will be used both as an adjective (operant behavior) and as a noun to designate the behavior defined by a given consequence.

A single instance in which a pigeon raises its head is a *response*. It is a bit of history which may be reported in any frame of reference we wish to use. The behavior called "raising the head," regardless of when specific instances occur, is an *operant*. It can be described, not as an accomplished act, but rather as a set of acts defined by the property of the height to which the head is raised. In this sense an operant is defined by an effect which may be specified in physical terms; the "cutoff" at a certain height is a property of behavior.

The term "learning" may profitably be saved in its traditional sense to describe the reassortment of responses in a complex situation. Terms for the process of stamping in may be borrowed from Pavlov's analysis of the conditioned reflex. Pavlov himself called all events which strengthened behavior "reinforcement" and all the resulting changes "conditioning." In the Pavlovian experiment, however, a reinforcer is paired with a *stimulus;* whereas in operant behavior it is contingent upon a *response*. Operant reinforcement is therefore a separate process and requires a separate analysis. In both cases, the strengthening of behavior which results from reinforcement is appropriately called "conditioning." In operant conditioning we "strengthen" an operant in the sense of making a response more probable or, in actual fact, more frequent. In Pavlovian or "respondent" conditioning we simply increase the magnitude of the response elicited by the conditioned stimulus and shorten the time which elapses between stimulus and response. (We note, incidentally, that these two cases exhaust the possibilities: an organism is conditioned when a reinforcer [1] accompanies another stimulus or [2] follows upon the organism's own behavior. Any event which does neither has no effect in changing a probability of response.) In the pigeon experiment, then, food is the *reinforcer* and presenting food when a response is emitted is the *reinforcement*. The *operant* is defined by the property upon which reinforcement is contingent—the height to which the head must be raised. The change in frequency with which the head is lifted to this height is the process of *operant conditioning*.

While we are awake, we act upon the environment constantly, and many of

the consequences of our actions are reinforcing. Through operant conditioning the environment builds the basic repertoire with which we keep our balance, walk, play games, handle instruments and tools, talk, write, sail a boat, drive a car, or fly a plane. A change in the environment—a new car, a new friend, a new field of interest, a new job, a new location—may find us unprepared, but our behavior usually adjusts quickly as we acquire new responses and discard old. We shall see in the following chapter that operant reinforcement does more than build a behavioral repertoire. It improves the efficiency of behavior and maintains behavior in strength long after acquisition or efficiency has ceased to be of interest.

57

Skinner Skinned

DANIEL C. DENNETT

Daniel C. Dennett (b. 1942) is currently Professor of Philosophy at Tufts University. He is the author of two recent books in philosophy of mind: Content and Consciousness *and* Brainstorms.

In "Skinner Skinned" Dennett offers a sympathetic, but ultimately quite critical analysis of Skinner's theory of behaviorism (reading 56). Dennett's first task is to try to fathom the reasons behind Skinner's theory. In particular, Dennett focuses on the question of why Skinner rejects explanations of human behavior which appeal to mental processes. Dennett calls these explanations, which refer to things like the agent's beliefs, desires, reasonings, or reflections, "intentional explanations," or sometimes "mentalistic explanations." According to Dennett, Skinner has a fairly reasonable objection to this type of explanation. The objection is that mentalistic explanations are too easy and they do not increase our understanding. For example, if a friend were to ask, "Why are you reading this book?" and you were to answer, "Because I want to," your friend might well feel that not much of an explanation had been provided. Still, Dennett maintains, against Skinner, that the beginnings of an explanation have been given. For your answer does rule out some possibilities, for example, that you are reading this book because you believe it will make you rich, or because you believe that reading is a good way to lose weight. Dennett believes that Skinner would be right only if mentalistic explanations had to stop at this level—with wants, desires, and so on. Then the explanations would be almost useless. Dennett locates Skinner's error in the belief that mentalistic explanations must terminate at this superficial level. If we deepen the explanation by explaining, for example, what a want is and where it comes from, then the mentalistic explanation can be viewed as the first step in a serious and illuminating theory.

B. F. SKINNER has recently retired, after a long and distinguished career at Harvard, and for better or for worse it appears that the school of psychology he founded, Skinnerian behaviorism, is simultaneously retiring from the academic limelight. Skinner's army of enemies would like to believe, no doubt, that his doctrines are succumbing at last to their barrage of criticism and invective, but of course science doesn't behave like that, and the reasons for the decline in influence of behaviorism are at best only indirectly tied to the many attempts at its "refutation." We could soften the blow for Skinner, perhaps, by putting the unwelcome message in terms he favors: psychologists just don't find behaviorism very *reinforcing* these days. Skinner might thing that was

unfair, but if he demanded *reasons,* if he asked his critics to *justify* their refusal to follow his lead, he would have to violate his own doctrines and methods. Those of us who are not Skinnerians, on the other hand, can without inconsistency plumb the inner thought processes, reasons, motives, decisions and beliefs of both Skinner and his critics, and try to extract from them an analysis of what is wrong with Skinnerian behaviorism and why.

. . .

Although counting myself among Skinner's opponents, I want to try to avoid the familiar brawl and do something diagnostic. I want to show *how* Skinner goes astray, through a series of all too common slight errors. He misapplies some perfectly good principles (principles, by the way, that his critics have often failed to recognize); he misdescribes crucial distinctions by lumping them all together; and he lets wishful thinking cloud his vision—a familiar enough failure. In particular, I want to show the falsehood of what I take to be Skinner's central philosophical claim, on which all the others rest, and which he apparently derives from his vision of psychology. The claim is that *behavioral science proves that people are not free, dignified, morally responsible agents.* It is this claim that secures what few links there are between Skinner's science and his politics. I want to show how Skinner arrives at this mistaken claim, and show how tempting in fact the path is. I would like to proceed by setting out with as much care as I can the steps of Skinner's argument for the claim, but that is impossible, since Skinner does not present arguments—at least, not wittingly. He has an ill-concealed disdain for arguments, a bias he feeds by supposing that brute facts will sweep away the most sophisticated arguments, and that the brute facts are on his side. His impatience with arguments does not, of course, prevent him from relying on arguments, it just prevents him from seeing that he is doing this—and it prevents him from seeing that his brute facts of behavior are not facts at all, but depend on an interpretation of the data which in turn depends on an argument, which, finally, is fallacious. To get this phantom—but utterly central—argument out in the open will take a bit of reconstruction.

The first step in Skinner's argument is to characterize his enemy, "mentalism". He has a strong gut intuition that the *traditional* way of talking about and explaining human behavior—in "mentalistic" terms of a person's beliefs, desires, ideas, hopes, fears, feelings, emotions—is somehow utterly disqualified. This way of talking, he believes, is disqualified in the sense that not only is it not science as it stands; it could not be turned into science or used in science; it is inimical to science, would *have* to be in conflict with *any* genuine science of human behavior. Now the first thing one must come to understand is this antipathy of Skinner's for all things "mentalistic". Once one understands the antipathy, it is easy enough to see the boundaries of Skinner's enemy territory.

Skinner gives so many different reasons for disqualifying mentalism that we may be sure he has failed to hit the nail on the head—but he does get close to an important truth, and we can help him to get closer. Being a frugal Yankee, Skinner is reluctant to part with *any* reason, however unconvincing, for being

against mentalism, but he does disassociate himself from some of the traditional arguments of behaviorists and other anti-mentalists at least to the extent of calling them relatively unimportant. For instance, perhaps the most ancient and familiar worry about mentalism is the suspicion that

(1) mental things must be made of *non-physical* stuff

thus raising the familiar and apparently fatal problems of Cartesian interactionism. Skinner presents this worry,[1] only to downplay it,[2] but when all else fails, he is happy to lean on it.[3] More explicitly, Skinner rejects the common behaviorist claim that it is

(2) the *privacy* of the mental

in contrast to the public objectivity of the data of behavior that makes the mental so abhorrent to science. "It would be foolish to deny the existence of that private world, but it is also foolish to assert that because it is private it is of a different nature from the world outside."[4] This concession to privacy is not all that it appears, however, for his concept of privacy is not the usual one encountered in the literature. Skinner does not even consider the possibility that one's mental life might be *in principle* private, *non-contingently* inaccessible. That is, he supposes without argument that the only sort of privacy envisaged is the sort that could someday be dispelled by poking around in the brain, and since "the skin is not that important as a boundary",[5] what it hides is nothing science will not be able to handle when the time comes. So Skinner suggests he will *not* object to the privacy of mental events, since their privacy would be no obstacle to science. At the same time Skinner often seeks to discredit explanations that appeal to some inner thing "we cannot see", which seems a contradiction.[6] For if we read these as objections to what we cannot

[1] *Beyond Freedom and Dignity* (New York: Knopf, 1971), p. 11. See also Skinner's *About Behaviorism* (New York: Random House, 1974): p. 31: "Almost all versions (of mentalism) contend that the mind is a non-physical space in which events obey non-physical laws".

[2] *Beyond Freedom and Dignity,* pp. 12 and 191.

[3] In the film, *Behavior Control: Freedom and Morality* (Open University Film Series). This is a conversation between Skinner and Geoffrey Warnock, reviewed by me in *Teaching Philosophy,* I, 2 (Fall, 1975): 175–7. See also *About Behaviorism,* p. 121: "By attempting to move human behavior into a world of non-physical dimensions, mentalistic or cognitivistic psychologists have cast the *basic* issues in insoluble form." Note that here he countenances no exceptions to the cognitivist-dualist equation.

[4] *Beyond Freedom and Dignity,* p. 191. See also Skinner's *Science and Human Behavior* (Free Press paperback edition, 1953): p. 285 and 82.

[5] "Behaviorism at Fifty", in T. W. Wann, ed., *Behaviorism and Phenomenology* (University of Chicago Press, 1964): 84.

[6] *Beyond Freedom and Dignity:* pp. 1, 14 and 193. In *About Behaviorism* Skinner countenances *covert* behavior (p. 26) and "private consequences" as reinforcers (p. 106), but on other pages insists "the environment stays where it is and where it has always been—outside the body" (p. 75), and "Neither the stimulus nor the response is ever *in* the body in any literal sense" (p. 148). See also "Why Look Inside", *About Behaviorism,* 165–69.

in principle see, to what is necessarily unobservable, then he must after all be appealing tacitly to a form of the privacy objection. But perhaps we should read these disparagements of appeals to what we cannot see merely as disparagements of appeals to what we cannot *now* see, but whose existence we are *inferring*. Skinner often inveighs against appealing to

(3) events whose occurrence "can only be inferred".[7]

Chomsky takes this to be Skinner's prime objection against mentalistic psychology,[8] but Skinner elsewhere is happy to note that "Science often talks about things it cannot see or measure"[9] so it cannot be that simple. It is not that all inferred entities or events are taboo, for Skinner himself on occasion explicitly infers the existence of such events; it must be a particular sort of inferred events. In particular,

(4) *internal* events

are decried, for they "have the effect of diverting attention from the external environment".[10] But if "the skin is not that important as a boundary", what can be wrong with internal events as such? No doubt Skinner finds *some* cause for suspicion in the mere internality of some processes; nothing else could explain his persistent ostrich-attitude towards physiological psychology.[11] But in his better moments he sees that there is nothing intrinsically wrong with inferring the existence of internal mediating events and processes—after all, he admits that some day physiology will describe the inner mechanisms that account for the relations between stimuli and responses, and he could hardly deny that in the meantime such inferences may illuminate the physiological investigations.[12] It must be only when the internal mediators are of a certain sort that

[7] *Beyond Freedom and Dignity*, p. 14.

[8] "The Case Against B. F. Skinner", *New York Review of Books* (December 30, 1971).

[9] "Behaviorism at Fifty", p. 84.

[10] *Beyond Freedom and Dignity*, p. 195; see also pp. 8 and 10. *About Behaviorism*, p. 18 and 170; *Cumulative Record* (1961): pp. 274–75.

[11] In "Operant Behavior", in W. K. Honig, ed., *Operant Behavior: Areas of Research and Application* (New York: Appleton Century Crofts, 1966), Skinner disparages theories that attempt to order the behavioral chaos by positing "some mental, physiological or merely conceptual inner system which by its nature is neither directly observed in nor accurately represented on any occasion by, the performance of an organism. There is no comparable inner system in an operant analysis" (p. 16). Here sheer internality is apparently the bogy. See also *Science and Human Behavior*, p. 32ff.

[12] He could hardly deny this, but he comes perilously close to it in *About Behaviorism*, where a particularly virulent attack of operationalism tempts him to challenge the credentials of such innocuous "scientific" concepts as the *tensile strength* of rope and the *viscosity* of fluids (pp. 165–66). Before philosophers scoff at this, they should remind themselves where psychologists caught this disease. A few pages later (p. 169) Skinner grants that a molecular explanation of viscosity is "a step forward" and so are physiological explanations of behavior. In "What is Psychotic Behavior?" (in *Cumulative Record*) he disparages "potential energy" and "magnetic field".

they are anathema. But what sort? Why, the "occult", "prescientific", "fictional" sort, the "*mental* way station" sort,[13] but these characterizations beg the question. So the first four reasons Skinner cites are all inconclusive or contradicted by Skinner himself. If there is something wrong with mentalistic talk, it is not necessarily because mentalism is dualism, that mentalism posits non-physical things, and it is not *just* that it involves internal, inferred, unobservable things, for he says or implies that there is nothing wrong with these features by themselves. If we are to go any further in characterizing Skinner's enemy we must read between the lines.[14]

In several places Skinner hints that what is bothering him is the *ease* with which mentalistic explanations can be concocted.[15] One *invents* whatever mental events one needs to "explain" the behavior in question. One falls back on the "miracle-working mind", which, just because it *is* miraculous, "explains nothing at all".[16] Now this is an ancient and honorable objection vividly characterized by Molière as the *virtus dormitiva*. The learned "doctor" in *Le Malade Imaginaire,* on being asked to explain what it was in the opium that put people to sleep, cites its *virtus dormitiva* or sleep-producing power. Leibniz similarly lampooned those who forged

> expressly occult qualities or faculties which they imagined to be like little demons or goblins capable of producing unceremoniously that which is demanded, just as if watches marked the hours by a certain horodeictic faculty without having need of wheels, or as if mills crushed grains by a fractive faculty without needing any thing resembling millstones.[17]

By seeming to offer an explanation, Skinner says, inventions of this sort "bring curiosity to an end". Now there can be no doubt that convicting a theory of relying on a *virtus dormitiva* is fatal to that theory, but getting the conviction is not always a simple matter—it often has been, though, in Twentieth Century psychology, and this may make Skinner complacent. Theories abounded in the early days of behaviorism which posited curiosity drives, the reduction of which explained why rats in mazes were curious; untapped reservoirs of aggressiveness to explain why animals were aggressive; and invisible, internal punishments and rewards that were postulated solely to account for the fact that unpunished, unrewarded animals sometimes refrained from or persisted in forms of behavior. But mentalistic explanations do not *seem* to cite

[13] *Beyond Freedom and Dignity,* pp. 9 and 23; *Cumulative Record,* pp. 283–84; "Behaviorism at Fifty".

[14] A patient and exhaustive review of these issues in Skinner's writings up to 1972 can be found in Russell Keat, "A Critical Examination of B. F. Skinner's Objections to Mentalism", *Behaviorism,* vol. I (Fall, 1972).

[15] "Behaviorism at Fifty", p. 80; *Beyond Freedom and Dignity,* Chapter 1, and p. 160.

[16] *Beyond Freedom and Dignity,* p. 195.

[17] *New Essays on the Understanding* (1704): Preface. See also Leibniz' *Discourse on Metaphysics,* X.

a *virtus dormitiva*. For instance, explaining Tom's presence on the uptown bus by citing his desire to go to Macy's and his belief that Macy's is uptown does not look like citing a *virtus dormitiva:* it is not as empty and question-begging as citing a special uptown-bus-affinity in him would be. Yet I think it is clear that Skinner does think that all mentalistic explanation is infected with the *virtus dormitiva*.[18] This is interesting, for it means that *mentalistic* explanations are on a par for Skinner with a lot of bad *behavioristic* theorizing, but since he offers no discernible defense of this claim, and since I think the claim is ultimately indefensible (as I hope to make clear shortly), I think we must look elsewhere for Skinner's best reason for being against mentalism.

There is a special case of the *virtus dormitiva,* in fact alluded to in the Leibniz passage I quoted, which is the key to Skinner's objection: sometimes the thing the desperate theoretician postulates takes the form of a little man in the machine, a *homunculus,* a demon or goblin as Leibniz says. Skinner often alludes to this fellow. "The function of the inner man is to provide an explanation which will not be explained in turn."[19] In fact, Skinner identifies this little man with the notion of an autonomous, free and dignified moral agent: he says we must abolish "the autonomous man—the inner man, the homunculus, the possessing demon, the man defended by the literature of freedom and dignity".[20] This is a typical case of Skinner's exasperating habit of running together into a single undifferentiated lump a number of distinct factors that are related. Here the concept of a moral agent is identified with the concept of a little man in the brain, which in turn is identified with the demons of yore. Skinner, then, sees superstition and demonology every time a claim is made on behalf of moral responsibility, and every time a theory seems to be utilizing a homunculus. It all looks the same to him: bad. Moreover, he lumps *this* pernicious bit of superstition (the moral-autonomous-homunculus-goblin) with all the lesser suspicions we have been examining; it turns out that "mental" means "internal" means "inferred" means "unobservable" means "private" means *"virtus dormitiva"* means "demons" means "superstition". Psychologists who study physiology (and hence look at *internal* things), or talk of *inferred* drives, or use mentalistic terms like "belief" are all a sorry lot for Skinner, scarcely distinguishable from folk who believe in witches, or, perish the thought, in the freedom and dignity of man. Skinner brands them all with what we might call guilt by free association. For instance, in *Beyond Freedom and Dignity,* after all Skinner's claims to disassociate himself from the lesser objections to mentalism, on p. 200 he lets all the sheep back into the fold:

> Science does not dehumanize man; it de-homunculizes him . . . Only by *dispossessing* him can we turn to the *real* causes of human behavior. Only then can we

[18] Skinner finds a passage in Newton to much the same effect as Leibniz: *Beyond Freedom and Dignity,* p. 9.

[19] *Ibid.,* p. 14.

[20] *Ibid.,* p. 200.

> turn from the *inferred* to the observed, from the miraculous to the natural, from the *inaccessible* to the manipulable. *(my italics)*[21]

But I was saying that hidden in this pile of dubious and inconsequential objections to mentalism is something important and true. What is it? It is that Skinner sees—or almost sees—that there is a special way that questions can be begged in psychology, and this way is *akin to* introducing a homunculus. Since psychology's task is to account for the intelligence or rationality of men and animals, it cannot fulfill its task if anywhere along the line it *presupposes* intelligence or rationality. Now introducing a homunculus does just that, as Skinner recognizes explicitly in "Behaviorism at Fifty":

> . . . the little man . . . was recently the hero of a television program called "Gateways to the Mind" . . . The viewer learned, from animated cartoons, that when a man's finger is pricked, electrical impulses resembling flashes of lightning run up the afferent nerves and appear on a television screen in the brain. The little man wakes up, sees the flashing screen, reaches out, and pulls the lever . . . More flashes of lightning go down the nerves to the muscles, which then contract, as the finger is pulled away from the threatening stimulus. *The behavior of the homunculus was, of course, not explained.* An explanation would presumably require another film. And it, in turn, another.*(my italics)* [22]

This "explanation" of our ability to respond to pin-pricks depends on the intelligence or rationality of the little man looking at the TV screen in the brain—and what does *his* intelligence depend on? Skinner sees clearly that introducing an unanalyzed homunculus is a dead end for psychology, and what he sees dimly is that a homunculus is hidden in effect in your explanation *whenever you use a certain vocabulary,* just because the use of that vocabulary, like the explicit introducton of a homunculus, presupposes intelligence or rationality. For instance, if I say that Tom is taking the uptown bus because he *wants* to go to Macy's and *believes* Macy's is uptown, my explanation of Tom's action *presupposes* Tom's intelligence, because if Tom weren't intelligent enough to put two and two together, as we say, he might fail to see that taking the uptown bus was a way of getting to Macy's. My explanation has a suppressed further premise: expanded it should read: Tom believes Macy's is uptown, and Tom wants to go to Macy's, so *since Tom is rational* Tom wants to go uptown, etc. Since I am relying on Tom's rationality to give me an explanation, it can hardly be an explanation of what makes Tom rational, even in part.

Whenever an explanation invokes the terms "want", "believe", "perceive", "think", "fear"—in short the "mentalistic" terms Skinner abhors—it must

[21] In *About Behaviorism,* (pp. 213–14) Skinner provides a marvelous list of the cognitivistic horrors—together with the hint that they are all equally bad, and that the use of one implicates one in the countenancing of all the others: " . . . sensations . . . intelligence . . . decisions . . . beliefs . . . a death instinct . . . sublimation . . . an id . . . a sense of shame . . . reaction formations . . . psychic energy . . . consciousness . . . mental illnesses . . . "

[22] "Behaviorism at Fifty", p. 80.

presuppose in some measure and fashion the rationality or intelligence of the entity being described. My favorite example of this is the chess-playing computer. There are now computer programs that can play a respectable game of chess. If you want to predict or explain the moves the computer makes you can do it mechanistically (either by talking about the opening and closing of logic gates, etc., or at a more fundamental physical level by talking about the effects of the electrical energy moving through the computer) or you can say, "If the computer *wants* to capture my bishop and *believes* I wouldn't trade my queen for his knight, then the computer will move his pawn forward one space," or something like that. We need not take seriously the claim that the computer *really* has beliefs and desires in order to use this way of reasoning. Such reasoning about the computer's "reasoning" may in fact enable you to predict the computer's behavior quite well (if the computer is well-programmed), and in a sense such reasoning can even explain the computer's behavior—we might say: "Oh, now I understand why the computer didn't move its rook."—but in another sense it doesn't explain the computer's behavior at all. What is awesome and baffling about a chess-playing computer is how a mere mechanical thing could be made to be so "smart". Suppose you were to ask the designer, "How did the computer 'figure out' that it should move its knight?" and he replied: "Simple; it recognized that its opponent couldn't counterattack without losing a rook." This would be highly unsatisfactory to us, for the question is, how was he able to make a computer that *recognized* anything in the first place? So long as our explanation still has "mentalistic" words like "recognize" and "figure out" and "want" and "believe" in it, it will presuppose the very set of capacities—whatever the capacities are that go to make up intelligence—it ought to be accounting for. And notice: this defect in the explanation need have nothing to do with postulating any non-physical, inner, private, inferred, unobservable events or processes, because it need not postulate any processes or events at all. The computer designer may know exactly what events are or are not going on inside the computer, or for that matter on its highly visible output device: in choosing to answer by talking of the computer's *reasons* for making the move it did, he is not asserting that there are any extra, strange, hidden processes going on; he is simply explaining the *rationale* of the program without telling us how it's done. Skinner comes very close to seeing this. He says:

> Nor can we escape. . . . by breaking the little man into pieces and dealing with his wishes, cognitions, motives, and so on, bit by bit. The objection is not that those things are mental but that they offer no real explanation and stand in the way of a more effective analysis.[23]

The upshot of this long and winding path through Skinner's various objections to mentalism is this: if we ignore the inconsistencies, clear away the red herrings, focus some of Skinner's vaguer comments, and put a few words in

[23] "Behaviorism at Fifty", p. 80.

his mouth, he comes up identifying the enemy as a certain class of terms—the "mentalistic" terms in his jargon—which when used in psychological theories "offer no real explanation" because using them is something like supposing there is a little man in the brain. Skinner never says the use of these terms presupposes rationality, but it does. Skinner also never gives us an exhaustive list of the mentalistic terms, or a definition of the class, but once again we can help him out. These terms, the use of which presupposes the rationality of the entity under investigation, are what philosophers call the *intentional idioms.*[24] They can be distinguished from other terms by several peculiarities of their logic, which is a more manageable way of distinguishing them than Skinner's. Thus, spruced up, Skinner's position becomes the following: *don't use intentional idioms in psychology.*

. . .

So let us put words in Skinner's mouth, and follow the phantom argument to its conclusion. We can, then, "agree" with Skinner when we read him between the lines to be asserting that no satisfactory, psychological theory can *rest* on any use of intentional idioms, for their use presupposes rationality, which is the very thing psychology is supposed to explain. So if there is progress in psychology, it will inevitably be, as Skinner suggests, in the direction of eliminating ultimate appeals to beliefs, desires, and other intentional items from out explanations. So far so good. But now Skinner appears to make an important misstep, for he seems to draw the further conclusion that *intentional idioms therefore have no legitimate place in any psychological theory.* But this has not been shown at all. There is no reason why intentional terms cannot be used provisionally in the effort to map out the functions of the behavior control system of men and animals, just so long as a way is found eventually to "cash them out" by designing a mechanism to function as specified. For example, we may not now be able to describe mechanically how to build a "belief store" for a man or animal, but if we specify how such a belief store must function, we can use the notion in a perfectly scientific way pending completion of its mechanical or physiological analysis. Mendelian genetics, for instance, thrived as a science for years with nothing more to feed on than the concept of a gene, a whatever-it-turns-out-to-be that functions as a transmitter of a heritable trait. All that is required by sound canons of scientific practice is that we not suppose or claim that we have reached an end to explanation in citing such a thing. Skinner, or rather phantom-Skinner, is wrong, then, to think it follows from the fact that psychology cannot make any *final appeal* to intentional items, that there can be no place for intentional idioms in psychology.

It is this misstep that leads Skinner into his most pervasive confusion. We have already seen that Skinner, unlike Quine, thinks that translation of intentional into non-intentional terms is possible. But if so, why can't intentional

[24] See, e.g., Roderick Chisholm, *Perceiving, A Philosophical Study* (1957), and numerous articles since then; also Quine, *Word and Object* (1960); W. G. Lycan, "On Intentionality and the Psychological", *American Philosophical Quarterly* (October, 1969).

explanations, in virtue of these bonds of translation, find a place in psychology? Skinner vacillates between saying they can and they can't, often within the space of a few pages.

. . .

In spite of his vacillation in print, it is clear that Skinner must come down in favor of the exclusive view, if his argument is to work. Certainly the majority of his remarks favor this view, and in fact it becomes quite explicit on p. 101 of *Beyond Freedom and Dignity* where Skinner distinguishes the "pre-scientific" (i.e., intentional) view of a person's behavior from the scientific view and goes on to say, "Neither view can be proved, but it is in the nature of scientific inquiry that the evidence should shift in favor of the second." Here we see Skinner going beyond the correct intuition that it is in the nature of scientific inquiry that ultimate appeals to intentional idioms must disappear as progress is made, to the bolder view that as this occurs intentional explanations will be rendered false, not reduced or translated into other terms.

I argue [elsewhere] that intentional and mechanistic or scientific explanations *can* co-exist, and have given [t]here an example supposed to confirm this: we know that there is a purely mechanistic explanation of the chess playing computer, and yet it is *not false* to say that the computer *figures out* or *recognizes* the best move, or that it *concludes* that its opponent cannot make a certain move, any more than it is false to say that a computer *adds* or *multiplies*. There has often been confusion on this score. It used to be popular to say, "A computer can't really think, of course; all it can do is add, subtract, multiply and divide." That leaves the way open to saying, "A computer can't really multiply, of course; all it can do is add numbers together very, very fast," and that must lead to the admission: "A computer cannot really add numbers, of course; all it can do is control the opening and closing of hundreds of tiny switches," which leads to: "A computer can't really control its switches, of course; it's simply at the mercy of the electrical currents pulsing through it." What this chain of claims adds up to "prove", obviously, is that computers are really pretty dull lumps of stuff—they can't do anything interesting at all. They can't really guide rockets to the moon, or make out paychecks, or beat human beings at chess, but of course they can do all that and more. What the computer programmer can do if we give him the chance is not *explain away* the illusion that the computer is doing these things, but *explain how* the computer truly is doing these things.

Skinner fails to see the distinction between explaining and explaining away. In this regard he is succumbing to the same confusion as those who suppose that since color can be explained in terms of the properties of atoms which are not colored, nothing is colored. Imagine the Skinner-style exclusion claim: "The American flag is *not* red, white and blue, but rather a collection of colorless atoms." Since Skinner fails to make this distinction, he is led to the exclusive view, the view that true scientific explanations will exclude true intentional explanations, and typically, though he asserts this, he offers no arguments for

it. Once again, however, with a little extrapolation we can see what perfectly good insights led Skinner to this error.

There are times when a mechanistic explanation obviously does exclude an intentional explanation. Wooldridge gives us a vivid example:

> When the time comes for egg laying the wasp *Sphex* builds a burrow for the purpose and seeks out a cricket which she stings in such a way as to paralyze but not kill it. She drags the cricket into her burrow, lays her eggs alongside, closes the burrow, then flies away, never to return. In due course, the eggs hatch and the wasp grubs feed off the paralyzed cricket, which has not decayed, having been kept in the wasp equivalent of deep freeze. To the human mind, such an elaborately organized and seemingly purposeful routine conveys a convincing flavor of logic and thoughtfulness—until more details are examined. For example, the wasp's routine is to bring the paralyzed cricket to the burrow, leave it on the threshold, go inside to see that all is well, emerge, and then drag the cricket in. If, while the wasp is inside making her preliminary inspection the cricket is moved a few inches away, the wasp, on emerging from the burrow, will bring the cricket back to the threshold, but not inside, and will then repeat the preparatory procedure of entering the burrow to see that everything is all right. If again the cricket is removed a few inches while the wasp is inside, once again the wasp will move the cricket up to the threshold and re-enter the burrow for a final check. The wasp never thinks of pulling the cricket straight in. On one occasion, this procedure was repeated forty times, always with the same result.[25]

In this case what we took at first to be a bit of intelligent behavior is unmasked. When we see how simple, rigid and mechanical it is, we realize that we were attributing too much to the wasp. Now Skinner's experimental life has been devoted to unmasking, over and over again, the behavior of pigeons and other lower animals. In "Behaviorism at Fifty" he gives an example almost as graphic as our wasp. Students watch a pigeon being conditioned to turn in a clockwise circle, and Skinner asks them to describe what they have observed. They all talk of the pigeon *expecting, hoping* for food, *feeling* this, *observing* that, and Skinner points out with glee that they have observed nothing of the kind; he has a simpler, more mechanical explanation of what has happened, and it *falsifies* the students' unfounded *inferences*. Since in this case explanation is unmasking or explaining away, it always is. Today pigeons, tomorrow the world. What Skinner fails to see is that it is not the fact that he has an explanation that unmasks the pretender after intelligence, but rather that his explanation is so simple. If Skinner had said to his students, "Aha! You think the pigeon is so smart, but here's how it learned to do its trick," and proceeded to inundate them with hundreds of pages of detailed explanation of highly complex inner mechanisms, their response would no doubt be that yes, the pigeon did seem, on his explanation, to be pretty smart.

[25] *The Machinery of the Brain* (New York: McGraw Hill, 1963): p. 82.

The fact that it is the simplicity of explanations that can render elaborate intentional explanations false is completely lost to Skinner for a very good reason: the only *well-formulated, testable* explanations Skinner and his colleagues have so far come up with have been, perforce, relatively simple, and deal with the relatively simple behavior controls of relatively simple animals. Since all the explanations he has so far come up with have been of the unmasking variety (pigeons, it turns out, do not have either freedom or dignity), Skinner might be forgiven for supposing that all explanations in psychology, including all explanations of human behavior, must be similarly unmasking.

It might, of course, turn out to be the case that all human behavior could be unmasked, that all signs of human cleverness are as illusory as the wasp's performance, but in spite of all Skinner's claims of triumph in explaining human behavior, his own testimony reveals this to be wishful thinking. Even if we were to leave unchallenged all the claims of operant conditioning of human beings in experimental situations,[26] there remain areas of human behavior that prove completely intractable to Skinner's mode of analysis. Not surprisingly, these are the areas of deliberate, intentional action. The persistently recalcitrant features of human behavior for the Skinnerians can be grouped under the headings of novelty and generality. The Skinnerian must explain all behavior by citing the subject's past history of similar stimuli and responses, so when someone behaves in a novel manner, there is a problem. Pigeons do not exhibit very interesting novel behavior, but human beings do. Suppose, to borrow one of Skinner's examples, I am held up and asked for my wallet.[27] This has never happened to me before, so the correct response cannot have been "reinforced" for me, yet I do the smart thing: I hand over my wallet. Why? The Skinnerian must claim that this is not truly novel behavior at all, but an instance of a *general sort* of behavior which has been previously conditioned. But what sort is it? Not only have I not been trained to hand over my wallet to men with guns, I have not been trained to empty my pockets for women with bombs, nor to turn over my possessions to armed entities. None of these things has ever happened to me before. I may never have been threatened before at all. Or more plausibly, it may well be that most often when I have been threatened in the past, the "reinforced" response was to *apologize* to the threatener for something I'd said. Obviously, though, when told, "Your money or your life!" I don't respond by saying, "I'm sorry. I take it all back." It is perfectly clear that what experience has taught me is that if I *want* to save my skin, and *believe* I am being threatened, I should do what I *believe* my threatener *wants* me to do. But of course Skinner cannot permit this intentional formulation at all, for

[26] But we shouldn't. See W. F. Brewer, "There is No Convincing Evidence for Operant or Classical Conditioning in Adult Humans", in W. B. Weimer, ed., *Cognition and the Symbolic Processes* (Hillsdale, New Jersey: Erlbaum, 1974).

[27] See *Science and Human Behavior,* p. 177, and Chomsky's amusing *reductio ad absurdum* of Skinner's analysis of "your money or your life" in his review of *Verbal Behavior,* in *Language* (1959), reprinted in J. Fodor and J. Katz, ed., *The Structure of Language, Readings in the Philosophy of Language* (New York: Prentice Hall, 1964).

in ascribing wants and beliefs it would presuppose my rationality. He must insist that the "threat stimuli" I now encounter (and these are not defined) are similar in some crucial but undescribed respect to some stimuli encountered in my past which were followed by responses of some sort similar to the one I now make, where the past responses were reinforced somehow by their consequences. But see what Skinner is doing here. He is positing an external *virtus dormitiva*. He has no record of any earlier experiences of this sort, but *infers* their existence, and moreover *endows* them with an automatically theory-satisfying quality: these postulated earlier experiences are claimed to resemble-in-whatever-is-the-crucial-respect the situation they must resemble for the Skinnerian explanation to work. Why do I hand over my wallet? Because I must have had in the past some experiences that reinforced wallet-handing-over behavior in circumstances like this.

. . .

I am suggesting that once Skinner turns from pigeons to people, his proffered "explanations" of human behavior are no better than this. If Skinner complains that mentalistic explanations are too easy, since we always know exactly what mental events to postulate to "explain" the behavior, the same can be said of all the explanation sketches of complex human behavior in Skinner's books. They offer not a shred of confirmation that Skinner's basic mode of explanation—in terms of reinforcement of operants—will prove fruitful in accounting for human behavior. It is hard to be sure, but Skinner even seems to realize this. He says at one point, "The instances of behavior cited in what follows are not offered as 'proof' of the interpretation", but he goes right on to say, "The proof is to be found in the basic analysis." But insofar as the "basic analysis" proves anything, it proves that people are not like pigeons, that Skinner's unmasking explanations will not be forthcoming. Certainly if we discovered that people only handed over their wallets to robbers after being conditioned to do this, and, moreover, continued to hand over their wallets after the robber had shown his gun was empty, or when the robber was flanked by policemen, we would have to admit that Skinner had unmasked the pretenders; human beings would be little better than pigeons or wasps, and we would have to agree that we had no freedom and dignity.

Skinner's increasing reliance, however, on a *virtus dormitiva* to "explain" complex human behavior is a measure of the difference between pigeons and persons, and hence is a measure of the distance between Skinner's premises and his conclusions. When Skinner speculates about the past history of reinforcement in a person in order to explain some current behavior, he is saying, in effect, "I don't know which of many possible equivalent series of events occurred, but one of them did, and that explains the occurrence of this behavior now." But what is the equivalence class Skinner is pointing to in every case? What do the wide variety of possible stimulus histories have in common? Skinner can't tell us in his vocabulary, but it is easy enough to say: the stimulus histories that belong to the equivalence class have in common the fact that they *had the effect of teaching the person that p,* of storing certain information. In the end Skinner

is playing the same game with his speculations as the cognitivist who speculates about internal representations of information. Skinner is simply relying on a more cumbersome vocabulary.

Skinner has failed to show that psychology without mentalism is either possible or—in his own work—actual, and so he has failed to explode the myths of freedom and dignity. Since that explosion was to have been his first shot in a proposed social revolution, its misfiring saves us the work of taking seriously his alternately dreary and terrifying proposals for improving the world.