

UNITS OF BEHAVIOR*

BERENT ENÇ†‡

*Department of Philosophy
University of Wisconsin-Madison*

1. Introduction. One major concern in cognitive psychology is explaining cognitively motivated behavior. However, attempts to specify the nature of the behavior that psychology is to explain have proved to be somewhat controversial. The aim of this paper is to conduct a preliminary investigation into the kinds of behavior psychology is (or should be) concerned with.

1.1 Background. Hornsby (1986) claims that there are two general approaches one can adopt conceiving the behavior of psychological agents. She paints two contrasting pictures to illustrate these approaches. The first picture shows the psychological subject as being affected by objects in the environment and affecting the environment. This involves one conception of behavior, according to which behavior is described in terms of the *effects* bodily movements have on the environment. According to this conception, we speak of an agent as, for example, having killed someone, drunk water, taken her umbrella—instead of speaking of her as having moved specific bodily limbs along specified trajectories. This is the conception that we employ in our common-sense discourse about human agents, and in that discourse we describe the explananda of our explanations in these terms.

The second picture is one in which descriptions of limb movements, or at least behaviors that are “close in” to the body, are adopted as the proper specifications of behavior, and internal states are seen to be the causal mediators between stimulus inputs and such limb movements. According to this picture, we describe the agent as having raised his arm, winked, walked, and so on.

Discussions in the contemporary literature tend to favor the second of

Received September 1994; revised January 1995.

†This essay has seen many revisions over the last ten years, and I am grateful to the graduate students and my colleagues at the University of Wisconsin-Madison for all their help. Special thanks for helping me to revise the most recent versions are due to an anonymous referee of this journal, and to Marc Bekoff, Eric Sidel, Larry Shapiro, Elliott Sober, and Dennis Stampe.

‡Send reprint requests to the author, Department of Philosophy, University of Wisconsin, 600 North Park Street, Madison, WI 53706.

Philosophy of Science, 62 (1995) pp. 523–542
Copyright © 1995 by the Philosophy of Science Association.

these pictures as the one that depicts the proper objects of psychological explanation. To pick just one clear expression of this wide-spread preference I quote from Kim 1982; the example he refers to in the passage is one in which he wants to turn on the stove to start cooking his dinner:

It is true that whether or not my action succeeds in bringing about the intended result normally depends on whether the belief involved is true. Thus, whether my action results in the burner being turned on depends on whether my belief that it would go on if the knob is turned is correct. However, it is not part of the object of *psychological* explanation to explain why the burner went on; all it needs to explain is why I turned the knob. It might be objected that not only did I perform the action of turning the knob but I also performed that of *turning on the burner*, and that this latter action does involve—it logically entails—the burner’s going on. That is correct; however, the action of turning on the burner, insofar as this is thought to involve the burner’s going on, is not an action that is the proper business of psychological theory to explain or predict. The job of psychological explanation is done once it has explained the bodily action of turning the knob; whether or not this action results in my also turning on the stove, my starting cooking dinner, my starting a house fire, and so on is dependent on events and facts quite outside the province of psychology, and are not the proper concern of psychological theory. Only *basic actions*, not ‘derivative’ or ‘generated’ actions, need to be explained by psychological theory. (Kim 1982, 64)

What Kim is maintaining here needs an argument to support it. And I suspect that the once-popular arguments in favor of psychological individualism may be adopted for this purpose. I intend to side step all that here. I will accept Kim’s suggestion that we take something like “basic actions” as the proper explananda of psychology. The purpose of this paper will be to offer several empirical considerations to help us to decide what exactly these “basic actions” should be. Kim’s motivation for offering basic actions as the proper explananda of psychology is derived from his conviction that psychological states that enter into psychological explanations are all supervenient on the internal states of the agent. Actually, the issue is best seen as a controversy over the extent of the supervenience base of psychological states rather than over the nature of psychological explanation.

I propose to show that the notion of basicness that is pertinent to behavior will not support Kim’s conviction. Perhaps I should clarify this a bit. Suppose, taking Kim’s definition of an internal property as a model (1982: 60), we define the notion of an *internal behavior* as follows: some behavior is internal if and only if its ascription to an agent does not entail

the existence of an object distinct from the agent (or the agent's body). So raising my arm, winking, tensing such and such muscles, and so on, are all internal, but turning on the burner, killing my assailant, taking the umbrella are non-internal. Now it is clear from Kim's concerns, and from the context in which the quoted passage occurs, that Kim takes the proper explananda of psychology to be purely *internal*. (Here I am assuming that his example of turning the knob, which is non-internal, because it presupposes the existence of a knob, is an inadvertent slip on Kim's part.) I will argue here that being internal is *not* an essential property of the proper explananda of psychology. And if this is true, then one set of arguments for the supervenience of the psychological over the physiological will have been undermined.

1.2 Summary of Thesis and Plan of Attack. The investigation here will be based on a conjecture that there exist certain units of behavior (what I shall later call "molar units") which are the simplest units for psychology—"atoms of behavior," so to speak, as far as psychological theorizing is concerned. These "atoms," i.e., the basic behavior that will be the candidate for the proper explananda of psychology, will turn out to be physiologically complex. Their relation to muscle movements and specific limb trajectories will mirror the relation of functionally conceived mental states to brain states. The first part of the thesis will be that psychological laws govern these functionally defined units rather than neuron firings and limb movements. In other words, this essay is a defense of functionalism in the domain of behavior.

My objective is to construct a careful argument for the existence such basic behavior in the animal kingdom. My working assumption is that philosophers' *a priori* attempts at refining our intuitions concerning the "basic acts" of human agents have not been very fruitful, and that more insight could be gained by studying the facts that are known about simple units of behavior in ethology.

The second part of the thesis will be that these units of basic behavior are not internal in the sense required by Kim.

The argument for the thesis that psychology requires a conception of basic behavior that cannot be defined exhaustively in terms of pure limb trajectories will contain two stages. The first stage will look at cases of learning from the perspective of a Skinnerian account of operant conditioning. My argument will aim at the preliminary conclusion that the Skinnerian account must suppose that organisms are equipped with the ability to directly produce certain functionally defined *units of behavior*, and that what is learned is when and under what circumstances to produce these units. A necessary condition for the mere possibility of the kind of learning that takes place in these simple organisms is that they be equipped with

such units. In the second stage I will look at some current physiological theories of the motor capacities of even simpler organisms, and I will argue that their findings support the view that such prepackaged units of behavior are part of the organisms' behavior repertoire.

2. Skinnerian Learning.

2.1 Introduction. Suppose we start with a pigeon who has been conditioned to peck at a key when a light goes on. Suppose, for example, the reinforcer is a grain of food that is mechanically introduced through a chute into the cage: when the key is depressed by sharp enough of an impact (sharp enough so that the device can discriminate between a peck and the pigeon's tripping, or its wing's pushing down, on the key), a micro-switch behind the key closes, and if the light is on, the circuit that operates the food chute also closes, and the grain drops down. So in this setup, the theory predicts that after the rigid reinforcement schedule has been applied, there will be a significant increase in the probability of the pigeon's pecking at the key when the light is on. When the probability does increase in this way, the light becomes a *discriminative stimulus* and the key pecking behavior becomes an *operant*.

An operant, as Skinner defines it, is a class of responses which "operates on the environment [of the organism] to generate consequences [for the organism]. The consequences define the properties with respect to which responses are called similar" (Skinner 1965, 65). The general idea is simply that an organism's behavior is shaped, frequencies in operants determined, *because* the behavior causes certain changes in the environment of the organism and these changes affect the organism in certain ways. Or to express the same idea, an input to the organism (a stimulus) elicits a response from the organism *because* in the past that type of response, when the same type of stimulus was present, has had certain consequences for the organism.

The force of these "because's" is clearly causal. (In saying this, I am departing from the letter of Skinnerian behaviorism) What we are here requiring is that a correlational fact (the correlation between key pecking under the stimulus conditions and food) become part of the cause of the key pecking behavior. When such a causal structure gets established, then the mere chance muscular movements that were occurring in the neck of the pigeon prior to the conditioning cease to be *chance* movements—these very movements are transformed from "pointless motion" into "behavior for some result." The difference between "pointless" motion and "behavior for some result" is not a difference in the type of *motion*, it is rather a difference in the causal ancestry of a type of motion. That is, whether a pigeon has pecked a key (so as to get food) or just moved its neck in a random way and accidentally happened to hit the key with its beak de-

depends on whether its neck movement was caused by some correlation between key pecking and food or not.

2.2 A Puzzle about Operant Conditioning. We can now ask an important question: What exactly is the pigeon conditioned to do? What *is* the behavior that we just said was behavior *for* some result? Is it moving the neck muscles in a specific way when the visual signals indicating the presence of a key-like object are received and processed? Is it pecking the key? Is it getting the key depressed? Is there a unique correct answer? These questions acquire greater weight when we remember that rats who learn to run a maze to get to the cheese, swim to the cheese when the maze is flooded, or crawl to the cheese when their legs are amputated. When we try to answer these questions in the context of operant conditioning, a problem emerges. To see this let us go back to our original observation about the causal structure involved in operant conditioning:

S: In operant conditioning, the fact that responses of a certain kind have in the past been correlated with certain consequences to the organism causes an increase in the probability of that kind of response.

In the context of Skinner's theory this is close to a definitional truth. But the uncertainty about the precise nature of the behavior involved raises in one's mind the possibility that the first occurrence of the word "response" in this claim need not refer to the same type of behavior as the second occurrence of that word.¹

In order to set up the problem clearly, we first need to ask, "what exactly is the mechanism that results in an increase in the probability of key pecking under stimulus conditions?" In spite of Skinner's abhorrence of causes, as I suggested above, it is hard to make sense of changes in probabilities unless something causes these changes. And I think the kind of causal story that is probably involved here is not a very unusual one. What we want is the past correlation between key peckings and food to become a causal factor in the causal link between the light flash and the pigeon's pecking the key. The process may be pictured this way: Before the conditioning, the light flash is not part of the cause of the pigeon's pecking the key. After the conditioning, the light has become the cause of the pigeon's pecking the key. What causes the light to be the cause of this behavior is just that past instances of this behavior have had certain desirable consequences to the pigeon. So the past correlation between key-pecking and food is a second order cause; it is that which causes the light to elicit (cause) a key-pecking behavior. The second order causation need

¹In a seminal essay, A. C. Purton (1978, 655–656) says, "the great defect of behaviourist psychology lies in its lack of attention to the task of classifying and making less ambiguous the use of the term 'behaviour'."

not be anything mysterious; it may indeed be reducible to the establishment of a particular state in the pigeon. After all, the fact that peckings have in the past been correlated with food may have induced some internal neuro-chemical state in the pigeon. And it may be the presence of this state that makes it possible (i.e., becomes a causal condition) for this flash now to cause this pecking. (Such a picture is developed in some detail by Dretske (1988) in his discussion of structuring causes.)

If this is the correct causal picture behind Skinnerian conditioning, the puzzle about the reference of the word “response” in *S*, above, may be stated more clearly now. Let us grant that the pigeon can peck that key in any one of a vast number of ways, that is let us grant that there is a vast number of sequences of muscular movements each of which is within the pigeon’s repertoire and each of which will trigger the mechanism that sends the food down the chute. Now consider this neurochemical state of the pigeon that was said to have been caused by the past correlation. A correlation between what and food is the cause here? Suppose the pigeon randomly produced 1000 sequences of muscular movements, a dozen of which got the key depressed. If the correlation is between each of these dozen movements and food, then we can’t expect the pigeon to peck the key in any way *other* than one of those twelve sequences. But it is a physiological fact that pigeons do produce *different* (new) sequences of muscular movements that get the key depressed *after* the conditioning. So what the pigeons learn is to do something that gets the key depressed. We may not know yet what the correct description of that thing is, but we do know that what they learn is *not* to exactly reproduce one of the dozen sequences that were used during the conditioning. So, and this is the mystery, when the light flashes and the pigeon pecks the key with a *new* sequence of muscular movements, how did each of those dozen sequences of muscular movements being followed by food increase the probability of this thirteenth *different* movement to be elicited by a light flash?

2.3 Provisional Solution of the Puzzle. It seems to me that there is only one satisfactory solution to this mystery. And that is to acknowledge that the proper conception of *response* in such conditioning experiments is a functionally characterizable “large” unit of behavior (perhaps *pecking at the key*), and not descriptions of specific neck movements (or of disjunctions of such movements).

There exist at least two possible scenarios, each of which can implement the causal process that underlies the conditioning:

(1) The “large” unit of behavior is somehow coded in the neural network of the animal. When a correlation between such a unit and, say, food, is established, this correlation induces a neurochemical state in the animal which makes it possible for the light flash to cause subsequent

tokenings of the unit of behavior. Here, since the unit comes “packaged” in the neurophysiology of the organism, the earlier responses that are correlated with food are *type identical* with the responses subsequent to the conditioning phase. The outward physical differences (i.e., differences in specific muscular movements and limb trajectories) among the different tokens of the same type of behavior are due either to random variations built into the neural circuits or to feedback responses to environmental inputs that are generated by the peripheral subsystems of the organism. These variations are not monitored by the higher centers of the nervous system, which mediate the conditioning process.

(2) The organism is able to “project” from the past correlation between *different* neck movements and food, and compute the type of behavior that will be required under the present circumstances. Here, the responses are all different, and the differences are bridged by some higher central computational system that “generalizes” from earlier experience by taking into account the differences in the circumstances and “hypothesizes” as to what to do next.

From a Skinnerian perspective, the differences between the two scenarios are immaterial. In each case, one defines the operant functionally and takes *that* to be the psychologically relevant unit of behavior. But from a neurophysiological point of view, the differences are obvious. I hope to show in the next section that empirical evidence points to the fact that the first scenario is quite real. As far as the second scenario is concerned, there is no denying that in the majority of cases of leaning experiments, the subjects do “project,” “generalize,” and compute which of the many *different behaviors* that are available to them will yield the distal goal (e.g., the cheese in the center of the maze). But I also hope to provide in the next section reasons to think that the “different behaviors” that are available to these subjects are not specific muscular movements or precisely describable limb trajectories; they are, rather, the “large units” referred to in the first scenario, each of which is defined functionally. In meeting these objectives, the next section will provide confirmation of the thesis that the “atoms” of behavior in psychology are functionally characterized units that encompass large variations in physiology. The thesis is actually in the spirit of Skinner’s account. His “operant” is a functionally defined unit: it is the *consequences* of the responses that determine the taxonomy of operants. So any response that opens the chute in the setup described will be a token of the same type of operant.

3. Physiology of Behavior.

3.1 Reflexes. Neuro-ethologists speak of two types of responses: *reflexes* and *motor programs*. A reflex is typically a “stimulus response circuit by which a particular input is inevitably followed, with little or no inter-

vening processing, by a unitary output” (Gould 1982, 72). Although the triggering of a reflex and its subsequent execution require very little contribution from the higher centers of the central nervous system, a reflex is a complex affair. Each muscle fiber is controlled by a single motor neuron. So a whole set of motor neurons need to fire in a patterned way for one whole muscle to be affected. Thus a reflex is constituted by a circuit that allows a specific type of input to generate a coordinated set of motor neuron firings that control a muscle or a group of muscles. The pathways included in a reflex circuit contain synaptic junctions which may, depending on the reflex involved, inhibit, facilitate, or chain the patterned firings of motor neurons that control the flexion or extension of the complete muscles.²

The complexity that exists in the architecture of the reflex circuits makes it possible for reflex behavior to exhibit a certain degree of plasticity. For example, the wiping reflex of the hind leg of the frog can apparently be triggered by applying a stimulus on the shoulder of the frog. The plots of the motion of the hind leg show a spectrum of trajectories. The variation from one complete wipe to another seems to be built into the circuitry and is not correlated with any external stimulus. In fact, neuro-ethologists conjecture that, even when external conditions do not vary, the system’s ability to generate a set of solutions for one and the same motor task may have been selected for because it provides the organism with a more effective strategy against unpredictable perturbations (Berkinblitt, Feldman, and Fukson 1986).

Perhaps more pertinent to my argument in this chapter are the types of plasticity of reflexes that arise in classical conditioning experiments, and are called “response generalization.” These experiments exploit reflexes that are naturally triggered by specific stimuli and through a process of conditioning, transfer the capacity to trigger the reflex to some other arbitrarily selected stimulus (the conditioned stimulus). One of the most fascinating cases of response generalization is reported by Wickens (1938). In this experiment a number of subjects had their arm strapped to a board with the palm downward. The subject’s middle finger rested on an electrode capable of delivering mildly painful shocks. Wickens conditioned an involuntary finger withdrawal response to a buzzer by giving several trials in which the buzzer preceded a shock to the finger. The response was involuntary in the sense that subjects could not inhibit the conditioned response when instructed to do so.

Here the same question that was asked about the pigeon’s key pecking response may be repeated: what is the response that is conditioned? Is it,

²Gallistel 1980 contains a summary of Sherrington’s pioneering work in the neurophysiology of reflexes. Reading Sherrington 1947, it is hard not to sympathize with him for thinking (erroneously) that all behavior is made up of reflexes chained together in complex ways.

for example, withdrawal from the surface? Or is it an excitation of the extensor muscles accompanied by an inhibition of the flexor muscles? The fact that there is a correct answer to these questions is revealed by the next stage of Wickens's experiment. When the involuntary response was thoroughly conditioned, Wickens unstrapped the subject's arm and then strapped it back onto the board again. Only now, the arm was turned over. The palm faced upward and the *back* of the middle finger rested on the electrode. Under these circumstances, the two descriptions of the subject's conditioned response—"withdrawal" versus "extensor excitation and flexor inhibition"—make opposite predictions about what will happen the next time the buzzer sounds.

In fact most (all except for one) of Wickens's subjects made a rapid withdrawal response the next time the buzzer sounded. In other words, the buzzer triggered *extensor inhibition* and *flexor excitation*. The nervous system somehow seems to know what to generate in order to withdraw the finger from the table.

One reason why this experiment is so intriguing is that it reveals the kind of intelligence we normally associate with agency to be already present in the "dumb" circuitry of a finger-jerk reflex. There is no higher center that "formulates hypotheses," or "generalizes from the data;" it is just that the reflex circuit is wired to move the finger *away from the surface* that normally produces the unconditioned stimulus. *That* is, as far as the reflex response to an electric shock is concerned, the "smallest" unit of behavior that the body is producing; individual subcomponents of that unit, i.e., individual muscle movements, are not independently accessible in the circumstances of this conditioning process.

3.2 Motor responses. The second type of response that has been the focus of ethologists' attention, i.e., a motor program, as a much more complex unit. These motor programs incorporate a highly structured series of interconnected neurons. They include not only circuits that coordinate a sequence of reflexes, but also oscillator circuits and circuits that are capable of servomechanistic control. These programs are triggered endogenously. When certain environmental signals are received, or when the internal chemistry of the system acquires the right hormonal characteristics (or both), a single "command neuron" fires and the whole network gets switched on. Examples of such motor programs that have been studied by ethologists make an impressive list. They include web-weaving of the female spider, claw-waving of the fiddler crab, copulatory behavior of the male praying mantis, bird songs, feeding behavior of the blowfly, and so on. An example involving the description of the cricket song might make the point clearer:

The wings are under the control of a set of muscles under the thorax

of the cricket. These muscles have connections to a network of nerve cells in the thoracic ganglion. Signals from some cells in the thoracic ganglion activate motor neurons, which in turn cause the wing muscles to contract in ways that result in a song. The trick is to get the motor neurons to fire as an integrated unit (because many muscle fibers are involved) with the appropriate timing pattern. This can be achieved by a single cell that runs from the brain in the head of the cricket to the thoracic ganglion. A neuroanatomist can find the *command interneuron* in its characteristic location in male cricket after male cricket. If a stimulating electrode is attached to a command interneuron, and the cell is stimulated by a weak electric current from the electrode, it produces a train of impulses with the correct timing. These signals reach neurons in the thoracic ganglion causing them to relay the right temporal pattern of orders to the wing muscles. The result is a perfect calling song, even in a brainless cricket. Presumably, in an intact cricket sitting by its burrow, cells in the brain integrate information about such things as the time of day and weather conditions. If the setting is correct, the brain cells fire. . . . (Alcock 1979, 37–38)

These motor programs are like reflexes in that when they are triggered, they are triggered as a whole package, and the neurophysiological subunits of the package are inaccessible to the triggering centers. However, one respect in which motor programs differ from reflexes is that while reflexes have external stimuli as their immediate causes, which trigger them directly, motor programs are *internally* triggered by the higher centers of the central nervous system. (For example, an immediate cause of the production of the cricket song is the output from the brain into the command interneuron.) These higher centers often integrate a set of sensory and proprioceptive signals before the triggering action takes place. A second respect in which the two types of behavioral output differ is derived from the high degree of complexity involved in motor programs. This complexity endows the output units with a plasticity that is of a higher order of magnitude than that of reflexes. Gallistel expresses this very well:

A cockroach's walking straight ahead may seem a simple and unitary act, but the close observer of cockroaches will note that they have a great many different ways of doing it. There are so many different leg-stepping patterns that it is only a slight exaggeration to say that the cockroach never walks in exactly the same way twice in its life. By "exactly the same way" I mean an exact duplication of the timing and magnitude of contraction and relaxation in every one of the many muscles involved, over several complete stepping cycles. (1981, 612)

These motor programs act as *units*, so that the higher centers of the animal have no access to the *components* of these units. This feature is

best revealed by Weiss's experiments on salamanders (1941). Weiss exchanged the right and left forelimbs of larval salamanders. When the limbs were re-innervated by regenerating nerves, the limbs resumed function. But since the adductor and abductor muscles, as well as the extensor and flexor muscles had traded places, the sequence of nerve signals which would normally produce forward locomotion, now ended up producing backward locomotion. So a hungry salamander when shown food ended up running away from it.

Some of these motor programs, like those of the salamander or the cockroach, come prewired and some develop upon maturation, but others are learned during the lifetime of the individual organism. The last category is the least well-known. But ethologists are fond of speculating. For example, Gould says:

Who thinks about which muscles to move when during walking, running, or crawling? It is as though these coordinated muscle patterns have become "hardwired," incorporated into the brain just as firmly as any wholly innate motor program. . . . And it seems clear that even motor tasks learned without any apparent innate guidance also become hardwired with time and practice: most of us can, after the months of arduous learning as children, tie our shoes with our eyes closed. (1982, 176)

3.3 A Preliminary Result. If we take stock of this brief and spotty survey of neuro-ethological literature, the following picture emerges: The electrochemistry of the cells of primitive organisms, or the central nervous system of higher animals equip them with units or patterns of behavior. Some of these units, like reflexes, are hardwired to be triggered externally and involve relatively linear paths. Others, like the motor programs, are more orchestrated patterns that are triggerable endogenously, and once triggered are played out in their entirety without the need for further endogenous control.

Going back to the opening question about the proper explananda of the theories of cognitive psychology, I suggest that an answer is to be found in the notion of a "basic" unit of behavior that is modelled on these motor programs. To avoid association with *a priori* arguments for or against basic acts, and to highlight the inner complexity of these units, I propose to label them, not "basic units," but "molar units of behavior." It will be assumed that in their architecture the molar units of behavior share the salient features of motor programs, i.e., their capacity to be triggered as a unit by the higher centers of the organism and a plasticity inherent to them that is unmonitored by the centers that trigger them.³ In

³Although many of the motor programs (or "fixed action patterns" as some ethologists

so doing, we still retain the idea that was central to basic acts: in our means-end computations what is “basic” is that behavior for the accomplishment of which no cognitively controlled behavior is a means. What makes molar units basic is that, in the particular circumstances surrounding their production, our cognitive centers relinquish control over all the things that we do with (or “in”) our bodies that constitute these units.

One immediate consequence of recognizing the existence of these pre-packaged large units of behavior is that it enables us to view the Skinnerian paradigm from a point of view slightly different from what Skinner suggests in his writings. When one thinks of operants as being built out of such units (as opposed to being constituted of physiologically describable responses), what appeared to be a mystery about operant conditioning becomes the expected routine. One now understands how a pigeon whose twelve *different* sequences of muscular movements “operated on its environment to generate of type of consequence for the organism” could produce a *new* sequence of muscular movements that generates the same type of consequence as a *result of* its past experiences. And the explanation lies in a simple shift. The shift requires us to move away from a picture where the centers that are involved in learning by operant conditioning are assumed to record and remember which specific muscle movements are correlated with the reinforcers. Instead, we adopt a picture where we view these centers as representing and controlling larger units of behavior (perhaps *pecking* in our example). We recognize that correlations are represented as holding between, say, food and the production of these larger units of behavior. Once we acknowledge that the features in virtue of which the muscular movement sequences are classifiable as *different* outputs are features that are not accessible to the centers involved in operant conditioning—that such differences arise from the role played by reflex circuits or servomechanistic devices that are impenetrable under the conditions to these centers—we see how the organism was literally and truly being reinforced to do *one and the same thing* each time it produced one of these “different” sets of muscular movements. On the other hand, it is clear that operants are not confined *only* to such motor programs. Organisms, like mice, can be conditioned to learn tasks (like getting to the cheese at the end of a maze) which may require them to recruit any one of several such programs (e.g., running, crawling when their legs are amputated, swimming when the maze is flooded). This fact was acknowledged above in the discussion above of the underlying causes of operant conditioning

call them) discussed above are *unmonitorable* by the higher centers, I do not intend their total impenetrability to be a necessary condition for these molar units of behavior. Just *the fact* that they are unmonitored (or the fact that they need not be monitored) *in the context* is sufficient for our purposes.

(scenario (2)). The point is rather that, as a matter of fact, operant conditioning rarely, if ever, works by reinforcing the organism to produce one specific set of muscular movements, because there is little evidence to think that specific muscular movement sequences are commonly under the control of the higher computational centers when learning by operant conditioning is in progress.

4. The Thesis that Molar Units are not Internal.

4.1 Constraints on Learning. At this point an interesting complication needs to be noted. The behaviorist ideal about complete control that seems to assume that *any* behavior can be reinforced and elicited by a discriminative stimulus turns out to be unrealizable. Literature on learning reveals that there are naturally imposed constraints on operant conditioning. In the experimental setup commonly described in the literature, reinforcers work most effectively if the behavior they are correlated with in the lab is the behavior the organism uses in its natural habitat in order to obtain that type of reinforcer. Experiments show how difficult it is to use reinforcers to get simple organisms to learn to produce molar units of behavior that are not naturally tied in their habitat to the reinforcers as means to ends. For example, Hineline and Rachlin (1969) report on their unsuccessful attempts to teach pigeons to peck a key where the reinforcer was the avoidance of an electric shock to their feet. All they got were several pigeons who learned to get the key depressed by beating it with their wings. Gallistel says in this context:

Shettleworth (1973, 1975) studied the effect of food reinforcement on six naturally occurring elements of golden hamster behavior—digging, scrabbling at the wall, rearing, face-washing, scratching with the hindfoot, and scent-marking. Food reinforcement greatly increased the frequency of the first three elements, digging, scrabbling and rearing, while having little or no effect on the frequency of the second three elements, face-washing, scratching and scent-marking. Her findings deliberately illustrate the point that was inadvertently illustrated by the work of Hineline and Rachlin: In a given conditioning context only certain elements in the lattice hierarchy are available for synthesis into a conditioned performance. Shettleworth's work eliminates the possibility that the failure of a given response to be conditioned by a given reinforcer may be due simply to the failure of that response to occur. One might object that keypecking did not easily become a conditioned escape response in Hineline and Rachlin's experiments because it rarely or never occurred and hence could not be reinforced. All of the behavioral elements that Shettleworth studied occurred with some frequency. Yet reinforcing some of them with food led to the

increase in frequency predicted with the “law of effect,” whereas reinforcing others with food did not. (1980, 314)

Marian and Keller Breland (1961) describe a series of cases where the animals who were being trained to perform certain tasks start “misbehaving.” One of the many examples they give is the unsuccessful attempt to teach pigs to pick up large wooden coins (“dollars”) and deposit them in a large “piggy bank.” This is how the Brelands describe the results:

At first the pig would eagerly pick up one dollar, carry it to the bank, run back, get another, carry it rapidly and neatly, and so on until the ratio was complete. Thereafter, over a period of weeks the behavior would become slower and slower. He might run over eagerly for each dollar, but on the way back, instead of carrying the dollar and depositing it simply and cleanly, he would repeatedly drop it, root it, drop it again, root it along the way, pick it up, toss it up in the air, drop it, root it some more, and so on.

We thought this behavior might simply be the dilly-dallying of an animal on a low drive. However, the behavior persisted and gained in strength in spite of a severely increased drive—he finally went through the ratios so slowly that he did not get enough to eat in the course of a day. . . . This problem behavior developed repeatedly in successive pigs. (1961, 683)

The authors name this phenomenon “instinctive drift.” Their explanation for it is the fact that behaviors that are strongly built into the species and are connected with their repertoire for gathering the reinforcing item in nature take over the manner in which they execute the task they are being taught to perform. The patterns to which the animals drift are “a violation of the so-called law of least effort. And most damaging of all, they stretch out the time required for reinforcement when nothing in the experimental setup requires them to do so. . . . Moreover, increasing the drive merely intensifies this effect” (Breland and Breland 1961, 684).

Two observations are in order here. The first is that these cases may be viewed as providing further support for the preliminary result presented above: in learning by conditioning, what sequence of muscular movements is being produced is not as relevant to the psychologists’ concern as is the identity of the molar unit that is involved. In showing this, we need to grant the possibility that the muscular movements of the limb constituting a token of scrabbling, say, are type identical to those that constitute a token of scratching. The difference between the two tokens will be due (among other things) to the position of the body relative to the limb, *and* to the result obtained through feedback. In other words, at least in those cases where the muscular movements involving the limb(s) in question are type identical, the difference between scrabbling and scratching will be

captured best by defining them functionally, i.e., in terms of the results they produce. Without any reference to such results, it would be hard to explain how one set of muscular movements (scrabbling) could be reinforced by food in one case, and the *same* set of muscular movements (scratching) could not be reinforced by food in the next case even when the hamsters were on increased drive. (Needless to say, on the Skinnerian paradigm, the need for such an explanation would never arise because the difference between physically identical scratching and scrabbings would never be detected, and the intransigent cases would have been swallowed up by the statistics.) On the other hand, if the types were defined by reference to the consequences, then we would have in this case two types, each encompassing variations in muscular movements, with the two sets of movements overlapping each other. When we then assume that the higher centers involved in learning control the *types* of behavior, not the specific movements of the limbs, the explanation of selective reinforcement would be obvious.

4.2 First argument for the Thesis: Teleology. The second observation one can make in these cases is the main reason for introducing them here. These cases highlight the fact that the units of behavior involved in such learning experiments are units that are *selected for specific tasks*. In other words, the functional characterization of these units has an element of goal-directedness built into it. Goal-directedness is a teleological notion, and the kind of teleology at issue here may be conceived etiologically. A favorite example of the use of teleology is found in the account of traits of animals by looking to see what these traits have been selected for (Sober 1984, Chapter 6). But the present observation needs no more than a very general picture of teleology. For our purposes, we can say that a trait (which can be a pattern of behavior) is a teleologically classified item if the fact that that trait has certain consequences to the organism explains why that trait is being exhibited by the organism (or is prevalent in the population to which the organism belongs). For example, one can explain why pecking is prevalent among birds by pointing out that pecking has a good chance of leading to food acquisition, and organisms which could exhibit this pattern of behavior were more fit than others (had a higher probability of having offspring and thereby could pass on their trait to the next generation). If *pecking* is classified as the trait that became prevalent in a population by such a process, then we can say that the class of pecking behavior has an essential element of teleology in it. To say this is to commit oneself to denying pecking status to any behavior which in its muscular movements is indistinguishable from a case of real pecking but which lacks the proper etiology (which was not made prevalent by the process described above). In this way, the distal causal ancestry of the

behavior becomes essential to its identity, and pecking thus ceases to be supervenient purely on the physiology of limb movements.

In order to see the essential role of such a teleological notion in the delineation of molar units, we need to go back to Skinner's definition of an operant. It will be remembered that Skinner says, "the consequences define the properties with respect to which responses are called similar." On a cursory reading of this claim, we might conclude that since, in the Shettleworth experiments, the setup makes the consequences of both digging and face-washing similar, these two behaviors should count as one type of operant. But the results reveal that doing so would distort the facts. (The computation in the change of the frequency of *that* operant would be confounded by the fact that although the frequency of digging goes up, that of face-washing does not.) The other alternative is to count as similar those responses *the function of which* is to generate the consequences in question.⁴ And what constitutes the function of a behavior is not determined solely by what consequences will, as a matter of fact, accrue to the organism in its present context.⁵ In the case of the golden hamster, in specifying the function of their behavior, one will need to advert to what that behavior was selected for in their natural environment. In other cases, internal representations that contain information about means-end relations may be needed in specifying what a function of a piece of behavior in a given context could be.⁶

If I am right in my claim that an element of teleology is needed in our functional characterization of some of these (molar) units of behavior, which, I suggest, are what Kim was looking for when he spoke of the proper explananda of psychology, then we have to give up the idea that the basic explananda of psychology are constituted by "internal" behavior. It will be remembered that an internal act was defined as an act the performance of which did not necessitate the existence of an object other than the agent. Face-washing, or rearing of the hamsters, for example, are not cases of purely internal behavior because these behaviors are not supervenient on just the physical movements of the hamsters.⁷ For the ham-

⁴There may be other alternatives here. I leave the search for the alternative that is least damaging to the Skinnerian program and is still consistent with the experimental results to the scholars of behaviorism.

⁵See Millikan (1984, 1989) and Neander 1991 for elaboration and defense of an etiological account of functions. Enç and Adams 1992 also provides a discussion of why a straightforward propensity account will not do in the analysis of functions.

⁶If the expression, "the function of a piece of behavior," is found to be infelicitous, some other appropriate teleological term, like "purpose," or "goal" may be substituted here.

⁷Ethologists who still find the "proximate-ultimate" dichotomy, which was suggested by Mayr (1961), useful are fond of defending the ultimate (or distal) descriptions of behavior, and in so doing, highlight the methodological importance of behaviors that are not fully supervenient on just the physical movements of the body. For a recent discussion, see Alcock and Sherman 1994.

sters' movements to count as that kind of behavior, they must have a particular history; they must have been selected for particular consequences. If teleology *is* an element in the characterization of these kinds of behavior, then the rearing in the golden hamster would be a different kind of behavior than rearing in some other species which may have been selected for fight. That is, if in this other species rearing was not developed because it provided an effective means for obtaining food, then for the purposes of psychological theory, it might be methodologically more correct to take the rearing in this organism to be a different piece of behavior from the rearing of the golden hamster. Perhaps the play bow of certain social carnivores illustrates better the point involved here. By adopting a certain posture, pups communicate to their playmates that what is to come is play and not aggression, predation or reproduction. The posture they adopt is called a "play bow" (Bekoff 1975; Bekoff and Allen 1992). The researchers have found remarkable physical similarities in the physical posture of the play bow across species. But it is clear that if some species adopted a different physical posture as the means for the same goal, for the ethologist it would still be a play bow. So a play bow is not only functionally defined, it also has an essential purposive element in its functional definition.

A critic may object here by arguing that just like "turning the burner on," "play bow" is an intentionally loaded description, and, as such, should be left outside the immediate scope of psychological theory. However, it is quite likely that there is nothing in the behavior of the dog that stands to *play bow* as *moving the hand in a certain way* stands to *turning the burner on*. Although the neurophysiological data are not yet available, it is quite conceivable that the dog's play bow is one prepackaged unit in much the same way as the motor program that generates the walking of a cockroach. If that is the case, then it may not be possible to redescribe it purely internally without doing serious methodological damage to the ethologists' program.⁸

I do not mean to defend the view here that all molar units of behavior are teleologically defined. Indeed, rearing and scratching are, as they stand, behaviors that may be fully supervenient on the limb movements of the organisms involved. What is crucial for my purposes here is that there is nothing in the nature and architecture of these units that makes it essential that they supervene on the limb movements and be internally defined. On the contrary, the empirical literature points to the importance of recognizing the adaptive value of these behaviors in their original definition. Once the behavior is classified in this fashion, due to homologies across species, it often turns out that the same type of internal movements

⁸This point is argued by Purton (1978) who suggests that there is no good way of defining *appeasement* across species as a cluster concept purely by its form.

is put into the service of the same type of task (e.g., the play bow, or scratching). In these cases, although the cluster of behaviors in question can be described fully internally, the original delineation for the class is teleologically motivated.⁹ Secondly, many higher organisms are capable of using one and the same behavior, which may originally have been selected for some specific task, as a means for various and unrelated tasks. Indeed the ability to learn to do novel things requires such a capacity. When that happens, the element of biologically selected purpose in the emergence of the behavior becomes secondary to the physiological description of the movement. For example, if rearing was originally directed in dogs at hunting and feeding, when it gets recruited for the purpose of begging at the table or for that of opening doors by leaning on the handle with front paws, it makes more sense to treat rearing as an internally defined behavior so that one can talk of the same behavior that is used for different ends. However, the “misbehaviors” discussed above show that teleological considerations grounded in natural selection are sometimes indispensable in classifying behavior for the purposes of operant conditioning. If *operant* is a theoretically viable tool in psychology, then recognizing the role of adaptation in the shaping of the behavior repertoire of the organisms renders readily explicable that which may otherwise have been puzzling.

4.3 Second Argument for the Thesis: Broad Functionalism. Totally distinct from considerations of teleology, there exists a second group of facts about motor programs that militate against the view that such units of behavior are always internal. When we look at descriptions of how a servomechanistic component of these programs functions, and how reflexes that are among the building blocks of these programs are triggered, we see that the shaping of these movements presupposes the existence of objects other than the body of the agent. The study of the cockroach’s walk makes clear how *the surface*, its inclination, and the pressure generated by the legs’ contact with the surface, are all essential to the successful execution of the walking pattern: What the cockroach is doing is *walking* (a non-internal piece of behavior), it is *not* moving its legs in the way it does when it walks under normal circumstances—there is no *one way* it moves its legs when it walks under normal circumstances. And I, too, am somewhat like the cockroach. I can tie my tie only when my fingers are in contact with it. When I try to go through the motions in the absence of a

⁹In an article titled, “Causal and Functional Questions: How Are They Linked?”, Curio (1994) argues, with rich examples from the study of the anti-predator behavior of birds, that purely internal (or “causal” as they are called in ethological literature) studies of such behavior remain incomplete when their function is ignored.

tie, I just cannot complete the mime. This suggests that my basic act here is not just the sequence of finger movements. It is rather tying a tie.

It might be objected here that the need for surfaces or objects in contact with the limbs is not essential to the architecture of these movements. The objection might go on to point out, quite rightly, that even if a cockroach cannot duplicate the walking motions of its legs when it is deprived of a surface, theoretically, a scientist could arrange to have the feedback stimuli administered in the proper sequence at the appropriate points of the nervous system of the cockroach, and thus induce in the cockroach an exact replica of the walking motion. Indeed, it also seems theoretically possible to compute all the variations in the amplitude and sequence of the signals that are consistent with *walking* and then simulate the right variety Galistel speaks of in the walk of a cockroach during its lifetime.

Correct as it is, what the objection ignores is the fact that computations of the scientist and the selection of the sequences of stimuli all presuppose the existence of a surface that would exert such and such pressures on the extremities of the limbs when the limbs produce such and such specific movements. The fact that one can, by studying the permissible variations in limb movements when they are in contact with a surface or an object, redefine a class of such movements fully internally, does not show that the class is an essentially non-internally defined class.

Finally, I think that it is a mistake to confine discussions of whether a psychological state or a behavior is internal or not merely to intuitions about the autonomy of psychology, about the role of the truth of one's beliefs in psychological explanation, about whether a clone of me has his psychological states identical to mine at the moment he is cloned or not. The considerations raised in this paper ought to make it clear that behavior and psychological states may fail to be internal in a much more modest way, in a way that has no direct implications to the latter fancy questions.

To summarize the point, then, the primary explananda of psychology need not be, and in general is not, if I may coin a phrase, behavior "narrowly conceived."

REFERENCES

- Alcock, J. (1979), *Animal Behavior*. Sunderland, MA: Sinauer Associates, Inc.
- Alcock, J. and Sherman, P. (1994), "The Utility of the Proximate-Ultimate Dichotomy in Ethology", *Ethology* 96: 58–62.
- Bekoff, M. (1975), "The Communication of Play Intention: Are Play Signals Functional?", *Semiotica* 15: 231–239.
- Bekoff, M. and Allen, C. (1992), "Intentional Icons", *Ethology* 91: 1–16.
- Birkinblitt, M. B., A. G. Feldman, and O. I. Fikson (1986), "Adaptability of Innate Motor Patterns and Motor Control Mechanisms", *Behavioral and Brain Sciences* 9: 585–638.
- Breland, K. and Breland, M. (1961), "The Misbehavior of Organisms", *American Psychologist* 16: 681–684.
- Curio, E. (1994), "Causal and Functional Questions: How Are They Linked?", *Animal Behaviour* 47: 999–1021.

- Dretske, F. (1988), *Explaining Behavior*. Cambridge, MA: MIT Press.
- Enç, B. and Adams, X. (1992), "Function and Goal Directedness", *Philosophy of Science* 59: 635–654.
- Gallistel, C. R. (1980), *The Organization of Action: A New Synthesis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- . (1981), Multiple book review of Gallistel (1980), *The Behavioral and Brain and Sciences* 4: 609–650.
- Gould, J. (1982), *Ethology: The Mechanisms and Evolution of Behavior*. New York: W. W. Norton & Company.
- Hineline, P. N. and Rachlin, H. (1969). "Escape and Avoidance of Shock by Pigeons Pecking a Key", *Journal of Experimental Analysis and Behavior* 12: 533–538.
- Hornsby, J. (1986). "Physicalistic Thinking and Conceptions of Behavior", in Pettit, P. and McDowell, J., eds., *Subject, Thought, and Content*. Oxford: Clarendon Press, pp. 95–115.
- Kim, J. (1982), "Psychophysical Supervenience", *Philosophical Studies* 41: 51–70.
- Mayr, E. (1961), "Cause and Effect in Biology", *Science* 134: 1501–1506.
- Millikan, R. (1984), *Language, Thought, and Other Biological Categories*. Cambridge, MA: The MIT Press.
- . (1989), "In Defense of Proper Functions", *Philosophy of Science* 56: 288–302.
- Neander, K. (1991), "The Teleological Notion of 'Function' ", *Australasian Journal of Philosophy* 69: 454–468.
- Purton, A. C. (1978), "Ethological Categories of Behaviour and some Consequences of their Conflation", *Animal Behaviour* 26: 653–670.
- Sherrington, C. S. (1947), *The Integrative Action of the Nervous System*. New Haven: Yale University Press. (First Edition, 1906).
- Skinner, B. F. (1965), *Science and Human Behavior*. New York: Macmillan Publishing Company, Free Press Paperback Edition. (First Edition, 1953).
- Sober, E. (1984), *The Nature of Selection*. Cambridge, MA: The MIT Press.
- Weiss, Paul (1941), "Self Differentiation of the Basic Patterns of Coordination", *Comparative Psychology Monographs* 17. Baltimore: Williams & Wilkins Co.
- Wickens, D. D. (1938), "The Transference of Conditioned Excitation and Conditioned Inhibition from one Muscle Group to the Antagonistic Group", *Journal of Experimental Psychology* 25: 127–140.