The Journal of Mind and Behavior
Winter 1992, Volume 13, Number 1
Pages 69–88
ISSN 0271–0137

The Physiology of Desire

Keith Butler

University of New Orleans

I argue, contrary to wide-spread opinion, that belief-desire psychology is likely to reduce smoothly to neuroscientific theory. I therefore reject P.M. Churchland's (1981) eliminativism and Fodor's (1976) nonreductive materialism. The case for this claim consists in an example reduction of the desire construct to a suitable construct in neuroscience. A brief account of the standard view of intertheoretic reduction is provided at the outset. An analysis of the desire construct in belief-desire psychology is then undertaken. Armed with these tools, the paper moves to an examination of the neural structures responsible for the production of motor behavior. This examination provides the basis for a theory of the neurophysiology of desire. A neurophysiological state is isolated and claimed to be type-identical to the state of desiring.

> The mathematician G.H. Hardy is supposed to have said a mathematician is someone who not only does not know what he is talking about but also does not care. Those who discuss in depth subjects such as the physiology of mind probably care, but I do not see how they could ever know. (Hubel, 1979, p. 39)

People commonly explain behavior, their own and others', by appeal to beliefs and desires, mental states to which we apparently have introspective access. But beliefs and desires can also be viewed as explanatory constructs, postulates of a powerful theoretical framework for the explanation of voluntary behavior. If this is right, then our justification for this so-called "belief-desire thesis" will depend on whether an appeal to beliefs and desires contributes to true explanations of voluntary behavior. Some materialists (P.M. Churchland, 1981; P.S. Churchland, 1986; Stich, 1983) have argued that a mature neuroscience is not likely to countenance such states in its explanations of behavior; call this view eliminativism. Others (Dretske, 1988;

I am indebted to Berent Enç, Robert Horton, Eric Saidel and Dennis Stampe for helpful discussions and/or comments on an earlier draft; I am also grateful for very helpful comments by J. Michael Russell and an anonymous referee of this journal. Requests for reprints should be sent to Keith Butler, Department of Philosophy, University of New Orleans, New Orleans, Louisiana 70148.

Fodor, 1976; Jackson and Pettit, 1990) have argued that it is impossible (logically, conceptually or metaphysically) that a belief-desire psychology should turn out to be radically false, and they (especially Fodor, 1976, 1987) have endeavored to demonstrate how one can be a materialist in the philosophy of psychology without endorsing eliminativism; call this view nonreductive materialism. (In introducing the term "materialism," I wish to make explicit my rejection of dualism. For the purposes of this paper, I will be assuming that dualism is false.) But eliminativists and the nonreductive materialists are united in the contention that psychological states (like beliefs and desires) will not reduce smoothly to neurophysiological states. They simply disagree about its implications; eliminativists think this nonreducibility calls for the elimination of belief-desire psychology, while nonreductive materialists think that it establishes the autonomy of belief-desire psychology as a special science. Whatever the implications of this nonreducibility claim, I believe that it is false. I think eliminativists and nonreductive materialists are both mistaken in claiming that belief-desire psychology is not reducible to a mature neurophysiology; eliminativists wrongly contend that belief-desire psychology will not survive the advance of neuroscience, while nonreductive materialists wrongly contend that belief-desire psychology is autonomous with respect to a mature neuroscience. I think the eliminativist is correct in demanding that the justification of belief-desire psychology depends on its reducibility to neurophysiology, and I think the nonreductive materialist is correct in retaining belief-desire psychology, but I will not argue for either of these claims in this paper. I propose in this essay to demonstrate that, as a matter of fact, belief-desire psychology is likely to reduce smoothly to neuroscientific theory. Belief-desire psychology is likely to survive the advance of neuroscience without being autonomous.

It will surely turn out that belief-desire psychology and neuroscience must *co-evolve* (in the sense described in Hooker, 1981, p. 49) if the reduction is to be smooth. In fact, from the examination of the role of desires in belief-desire explanations, we will see that not everything that the poets and songwriters count as desires will be accommodated by the proposed reduction. I do expect, however, to repay the debt. The proposed reduction will (as reductions do) explain certain features of the reduced concept that otherwise go unexplained.

I will be limited in this paper to an examination of the reducibility of desire. My hope is simply that if the state of desiring can be shown to be reducible to an appropriate neuroscientific state type, then a similar program might be executed for belief. I will begin with a brief account of the standard sense of theoretical reduction. I will then turn to an analysis of the role of desires in belief-desire explanations of voluntary behavior, focusing in particular on how desires function differently from beliefs in

such explanations. Belief-desire explanations are causal explanations, so a successful account of the explanatory role of desires must explain how desires contribute to the causation of movement and action. Such an account must, or so I will argue, treat desires physiologically, since it is, in the end, certain physiological states that move us to action.[1] I will develop a physiological picture within which a role for desire will be cast. Desires, I will maintain, arise in response to particular aspects of imbalance in the internal environment of an organism, and move the organism to action as a result.

Russell (1984) has argued that belief-desire explanations are not causal. Briefly, the main reason for Russell's claim is that ordinary language reports of actions do not allow for them to have been "brought about" by anything. Causal explanations of actions, however, essentially involve the presumption that actions are "brought about" by something; hence they cannot be correct. I do not agree with this line of reasoning mainly because I view causal explanations of movements as providing an essential ingredient in the analysis of actions. I grant that actions are not *just* caused movements, but I contend that they are *at least* caused movements; hence the explanation of an action essentially involves appeal to the cause of the movements involved in the action. The belief-desire thesis just is that the causes of these movements are beliefs, desires and the like.

This is a large issue that would take us too far afield should we devote to it the attention it deserves. I wish only to acknowledge explicitly an opposing position to the one I adopt and to give some idea why I reject it.

Reduction

We should be clear straight off about how to understand the sense of reduction at issue. For the sake of the argument we will assume that theories that enter into the reduction relation are sets of sentences, usually of the form of universal generalizations, and that the reduction relation is assymetric. Our understanding will then be this:

> i) A theory A reduces to a theory B when and only when a given nonlogical predicate found in the laws and generalizations of theory A can be linked via "bridge laws" to a corresponding nonlogical predicate found in the laws and generalizations of theory B; and
>
> ii) All consequences of theory A are consequences of theory B plus the bridge laws.

---

[1] This type of analysis has not, so far as I know, received adequate attention in the literature. Clark (1980), for example, discusses only a very limited domain of psychological constructs, and nothing like the folk psychological constructs with which I shall deal.

The issue, then, is whether the predicates in theory A refer to properties that are *type-identical* to the properties referred to by the appropriate predicates of theory B. In other words, reduction requires that the predicates of theory A be *coextensive* with the appropriate predicates of theory B. The importance of this sense of reduction is that the laws, generalizations and explanations framed in terms of the reduced theory can instead be framed in terms of the reducing theory; the laws and generalizations of the reduced theory can be derived, via appropriate bridge principles, from the reducing theory. The reduced theory then becomes dispensable, though it may be retained in the way that thermodynamics has been retained despite its reduction to the kinetic theory of gases; in such cases, the reduced theory is simply shown to be a (more or less accurate) fragment of a larger theory. One must be careful to notice that i and ii above do not imply that theory B is reducible to theory A; no claim is made that there are bridge laws connecting all the predicates in theory B with the predicates of theory A. The reducing theory is in some sense larger than the reduced theory; it explains and/or describes more.

The position I will be advancing in this essay, then, is that the predicates of belief-desire psychology will reduce to the predicates of a mature neuroscience in just this sense. My case for this position will simply be an example reduction of the desire construct to a certain type of neurophysiological state. The aim is not to eliminate the concept of desire in favor of some neurophysiological concept anymore than the aim of the reduction of temperature to mean kinetic energy eliminates the concept of heat. In other words, one can be a reductionist without being an eliminativist. That, at any rate, is the sort of picture I will be drawing.

Desire vs. Belief

A central question in the analysis of desire is how desires differ from beliefs, in particular, beliefs about what would be good to do or have, and beliefs about how to get what we want. We apparently do not want everything that seems good to us, nor everything that is a means to what we want; the fact that jogging, for example, seems to be a good thing does not mean that I want to do it. That it seems good to me may just reflect the belief I have that jogging is a means to something that I do want, or it may just be something that seems good, but not, as it were, good enough to want.

That desires differ from beliefs is perhaps best revealed by an examination of the role each plays in the orthodox practical syllogism:

Desire (X)
Believe (Y will get X)

Do, or Intend to Do, (Y).

It has been known, I suppose since Aristotle,[2] that a practical syllogism consisting solely of beliefs would not adequately explain an organism's behavior. One could believe X would be good, and believe that doing Y would result in the acquisition of X, but still not do Y. The reason for this, of course, is that in such a case one lacks the *desire* for X; mere belief is insufficient to cause voluntary behavior. By the same token, of course, a practical syllogism consisting solely of desires would not be adequate to the explanatory task either. One could desire X, but without some knowledge of the means through which one must go in reaching X, one could never execute behavior designed for that purpose. At best we would get chaotic flailings with little hope of reaching the object or state of affairs desired. Beliefs, it would seem, inform us about the world, tell us how to get what we want, but it is desires that move us to action. Expression of a similar sentiment can be found in Stampe (1988, section 7) and, again, Aristotle.

This, I take it, is the conviction behind a belief-desire psychology. One state type is responsible for the acquisition of information about the world, the other is responsible for the initiation of behavior. I do not, of course, mean to erect an impermeable barrier between these two state types, for surely they conspire; desires determine to a large extent the parts of the environment about which we become informed, and beliefs modify the behavior we produce in navigating through the environment. This interaction, however, should not obscure the very different roles each is thought to play in the production of voluntary behavior.

Beliefs are *cognitive* states, states that deliver information about the world. Knowing the means to the object of desire, for example, is a piece of cognition. Desires, on the other hand, are *conative* states, states that impel the organism to action. It is this impetus to act that is missing in a psychological framework consisting solely of beliefs. This, I submit, is essential to the nature of desire, for this is the mark by which desires are distinguished from beliefs.

But, in addition to their initiative capacity, desires also have *objects*, something they are *for*. Thus, they differ from beliefs as well in that they provide the end or consequence at which the behavior is directed. No belief of this sort, a belief that specifies the goal of behavior, is needed in the practical syllogism. The contrast here is that it is essential to the explanation of behavior that desires have this object (because there is something the desire is for), whereas any belief that specifies the goal of the behavior is not essential to the explanation of that behavior. Desires do not merely cause behavior, they

[2] See, for example, *De Anima* III. 9 432a15: "The soul of animals is characterized by two faculties, (a) the faculty of discrimination which is the work of thought and sense, and (b) the faculty of originating local motion."

cause behavior that ideally has a particular result. This result, the object of desire, must, from an explanatory perspective, be what it is the acquisition of which would (*ceteris paribus*) terminate the behavior that the desire has caused. For the complete description of the explananda of folk psychology would include that the behavior has some certain result and terminates upon the achievement of the result: desires move one to action that will ideally satisfy that desire. Thus, whatever plays the explanatory role of desire must fix which consequence of the behavior will terminate that behavior, which result will satisfy the desire. That is the explanatory function of the *object* of desire. For this reason we may be confident that the object of desire is what *will* satisfy the agent, and hence terminate the relevant behavior, not what the agent *thinks* will satisfy her. Of course, in cases where the agent correctly thinks that an object x will satisfy her, it will be trivially true that what the agent thinks will satisfy her is the object of desire. My point is only that in cases where what will actually satisfy her and what she thinks will satisfy her diverge, it is what will actually satisfy her that is the real object of desire. This, I take it, is just part of our ordinary concept of desire; and this is one aspect of that concept that the reductive analysis will vindicate. What the agent thinks will satisfy her may in fact not, in which case (*ceteris paribus*) the agent will not have lost the impetus to action and will presumably seek other means to the satisfaction of her desire. One may, of course, imagine deviant cases where an agent acquires what she wrongly thinks will satisfy her desire, but where the impetus to action is quieted nonetheless simply because she thinks that she has acquired the object of her desire. The present thesis can accommodate such cases in either of two ways: (a) the behavior triggered by the relevant desire is aborted before it has reached the real object of desire (i.e., the *ceteris paribus* clause is violated), or (b) the object of desire somehow includes what is thought to be its satisfaction condition, so that the agent will be satisfied with thinking that she has achieved what she wanted to achieve, even if, in some *other* sense, she really has not.

Let us grant, then, that desires have the two-fold character of determining the beginning *and* end of voluntary behavior, and that this is a function of the explanatory structure of belief-desire explanations. It may be the domain of our cognitive mechanisms to determine the means through which we reach that end from this beginning, but it is the domain of our conative mechanisms to begin voluntary behavior and specify its proper end.

This is, in a sense, just as it should be. What better way to design a system than to produce, only when it has to, costly energy expending behavior, behavior that is directed toward the fulfillment of the need from which it arose? From an evolutionary perspective it is no wonder conative states impose direction on the behavior they cause; they have been selected because they do just that. That is what they do that is of benefit to the

organism. The picture that this suggests, of course, is that the object of desire is to be fixed by the biological function of our conative mechanisms, and how it is that those mechanisms serve this function. This deserves a closer look.

Several views in the literature on cognitive states hold that the object of belief is given by a causal relation between some state of the environment and our cognitive mechanisms. A particular state of our cognitive mechanisms is said to *represent* some state of the environment if our cognitive mechanisms were caused to be in that state by the state of the environment (Dretske, 1988; Matthen, 1988; Stampe, 1977). Since there are indefinitely many states of affairs that can and do enter into the causal relations with our cognitive mechanisms, we must, if we are to capture the intentionality of *belief*, specify only a certain subset of these states of affairs as (potential) objects of belief. For this reason we isolate the *function* of the cognitive mechanisms, what it is they do that is of some benefit to the organism.

There are a variety of ways of assigning functions. Some appeal to natural selection (see Wright, 1973), while others appeal to learning (see again Dretske, 1988). The reader is invited to consider as well Millikan (1984). I do not here commit myself to any particular theory of function attribution; I only wish to cite with approval the practice of assigning the objects of belief and desire based on an appeal to the function of the mechanisms involved, however those functions are properly specified. I have attempted in this paper to state the basis of function attribution in as neutral a fashion as possible for just this reason.

When these cognitive mechanisms are performing their function, their causal relations with the environment determine the content of the belief to which they give rise. This approach dates back at least to Helmholtz (1962, p. 2), and several problems for it have been raised in the literature (Dennett, 1987, pp. 301–307; Fodor, 1987, chapter 4). I do not think these problems are insuperable; at any rate, it is not my purpose to produce an analysis of belief, so I will not pursue this any further (though see Forster, 1987, for an impressive treatment of these problems). I mention this approach to an understanding of belief only to place the present analysis of desire in the proper context.

If there is virtue in the thesis that cognitive mechanisms acquire their objects by having the function of *being caused* to be a certain way by some state of the environment, then perhaps there is virtue to the thesis that desires acquire their objects by having the function of *causing* certain states of affairs (see Stampe, 1986; Searle, 1983, chapter 1). As with belief, there is the problem of providing a principled reason to isolate some certain link in the causal chain as the object of the psychological state, in this case, desire.

David Papineau has attempted a solution to just this problem.

> An action stemming from a desire will have a concertina of effects which are relevant to its enhancing inclusive fitness. As we proceed outwards, so to speak, we will go past

> effects which are taken to be relevant only in virtue of current beliefs, to effects the relevance of which is assumed by natural selection but not by the agent, and ending up (if everything works well) with enhanced fitness. The satisfaction condition of desire is the first effect which is taken to be relevant by evolution but not by the agent. That's what the desire is *for*—to give rise to actions which have *that* effect. It's not for earlier effects, because whether actions have those effects at all depends on what beliefs the desire is interacting with. And it's not, in the first instance, for later effects, because the actions it directs are designed to produce those later effects only *through* producing that first effect. (Papineau, 1984, p. 564)

Contained herein is a pair of principles according to which the object of desire can be isolated from the rest of the causal chain to which it belongs.

> 1. The object of desire does not depend on beliefs with which the desire is interacting.
>
> 2. The object of desire is the element in the causal chain through which desires have been selected to cause continued survival.

Desire-caused eating behavior might result in the acquisition of a spoon, the removal of ice cream from the refrigerator, the eating of the ice cream, nourishment, the restoration of internal balance and continued survival. According to Papineau, it is the eating of the ice cream that is the object of desire since that is the only one of these events that satisfies the pair of conditions above. As a heuristic, one might imagine that when our actions intersect with the environment (as they invariably do), it is events at that intersection that are the objects of desire. Events prior to that intersection are mere means to the satisfaction of desire, and events after it are mere consequences that nature has assumed would accrue to the behavior. What we want is to eat ice cream just as our folk wisdom would have it.

There is an apparent and immediate plausibility to this tandem of principles. The first principle demonstrates why a belief-desire psychology need not, and in fact does not, commit one to the view that one must desire the means to the satisfaction of one's desires. This is, as Papineau notes, the domain of our cognitive mechanisms, not our conative mechanisms. We do plenty of activities, for example, taking out the garbage, that we simply do not want to do. Desire generated action results in this only because of the beliefs with which it is interacting. We take out the garbage because we *believe* that it is a means to, perhaps, freedom from malodor.

The second principle offers an account of another dimension of the intentionality of desiderative states. We describe ourselves as being, often enough, overcome with a desire for ice cream-eating, but not for the bloating and nausea that inevitably follow such an event. The object of the desire does not include the bloating and nausea because the action performed that eventually results in it does so only by first causing the ingestion of the ice cream. The same goes, of course, for the evolutionary consequences to which Papineau was addressing himself.

I now want to develop a physiologically mature conception of desire, building on what is known about the nature of motor control. I hope to present a compelling case that the concept of desire that we have just sketched will find a home in the emerging physiological understanding of the origins of voluntary behavior. I will be working from Gallistel's (1980) discussion of motor behavior. I will suggest how the basic ideas found there can be extended beyond the limited domain of motor control to the rich, relatively more plastic domain of "intentional" behavior.

The Organization of Behavior

Current research on the organization of the neural structures that cause voluntary movement stresses that the initial neural command to execute a movement does not specify the details of its implementation (see Ghez, 1985; Ghez and Fahn, 1985); the structures that cause motor output are arranged hierarchically. Consider the following analogy. An army general does not stipulate how each troop is to behave at each moment when he hands down the order to destroy the next village; he leaves that to the colonels and captains. The colonels and captains similarly do not determine the specific behavior of each soldier; that is left to the ingenuity of the individual. Nevertheless, the collective efforts of all these people do ideally result in the massacre the general ordered; the massacre was caused by, and can be explained by appeal to, the general's order. A general's order, then, can determine the ultimate outcome without determining the method by which that outcome comes to pass. Because of this, a type identical order can result in a type identical massacre without there being any relevant type identity among the specific means to this result.

A very similar hierarchical control structure is in operation in the production of human motor behavior. A circuit at the highest level sends a certain "command" signal, or instruction, to a set of circuits whose domain of control is somewhat smaller. The circuits receiving this signal can, depending upon the prevailing circumstances, execute this task in various ways. The signal it receives, however, does not determine which of these ways will in fact be realized; it determines at most that one of them will. Which one actually is activated depends on contingent conditions of this circuit and other signals impinging upon it. This kind of variable implementation of an instruction can filter down through the neural hierarchy indefinitely, resulting in an indefinitely wide variety of realizations of the original instruction. At some point the signals will reach a level of circuits whose job it is to activate motor neurons and produce contractions and extensions of muscle fibers. The pattern of excitation depends entirely on the nature and prevailing conditions of the control hierarchy. A single type of command can have

a vast array of possible realizations; and each realization will be a realization of that command. They will all have the same outcome (provided nothing external or unusual interferes) since they will all be guided by the same instruction.

When a signal is sent from, say, the level[3] of whole muscle activation, to the level of the motor unit, it is not the case that a set of circuits at level of the motor unit is activated; rather, a certain larger set of circuits at that level are, in Gallistel's terminology, *potentiated*. That is, the circuits in this larger set at level 1 are not automatically activated when they receive a signal from the circuit at level 2; they are merely put in a state wherein the probability of their being activated is greater than it was previously. At the same time, another set of circuits at level 1 is *depotentiated*; the probability of their being activated is lowered. The reason for this selective potentiation and depotentiation is that the effectiveness of a certain pattern of neural excitation may (and virtually always does) depend on certain circuits not being active. The pattern of excitation appropriate for sending me forward, for example, cannot be effective if crucial elements of the pattern appropriate for sending me backward are also active. But which particular circuits are activated depends on certain other conditions (to be discussed shortly). It is the principle of selective potentiation and depotentiation according to which a cause can determine a distal effect without determining the proximal effects by means of which the cause determines the distal effect. On a relatively microcosmic level, for example, a cause, such as a circuit at level 2, potentiates a set of circuits at level 1, each of which, under different circumstances, will lead (perhaps by potentiating another lower set of circuits) to the distal effect. It is this selective potentiation and depotentiation that is the feature of neural interaction that allows causes to determine distal effects without determining the proximal effects by means of which the distal effect is produced.

I have alluded to the fact that whether a certain circuit is activated depends on whether certain further conditions obtain. This must be explained. Whether or not it is appropriate to move my middle finger in grabbing the cheese or my index finger depends, for example, on the health and relative filth of the fingers involved. It is, however, not essential to the mere decision to grab the cheese that I also decide which finger I will use. Which particular muscle fibers will contract within the chosen fingers depends on the size and shape of the cheese, and the orientation of my hand to it. Nevertheless, all of the motor neurons controlling fibers relevant to the grabbing have been potentiated. It is just that only certain of them will be called into action. And the manner in which they do get activated bears a striking resemblance

[3] For a vividly detailed account of the different levels in the structure of this hierarchy see Gallistel (1980, pp. 275–280).

to the manner in which my arm, or even my whole body, would get activated. To explicate this I must introduce the notion of a *servomechanism* (on which see also Merton, 1973).

A servomechanism is a type of control unit that takes a sensory input signal representing a certain aspect of the current state of affairs and matches it to a stored representation, called an *efference copy*. Any discrepancy between the two representations results in an error signal being issued that has the function of potentiating or activating whatever muscle groups or motor neurons are necessary in order to eliminate the error signal, or secure an isomorphism between the two representations. This is the principle according to which a guided missile adjusts to the target onto which it is locked, and the principle according to which a potentiated circuit gets activated. Which muscle groups get activated in my hand when it is grabbing for the cheese depends in part on which ones are required to bring it about that the higher instruction (getting the cheese) is executed. Many of them are potentiated, but only those that will lock onto the cheese will be activated. Thus, if my thumb and index finger are necessary to the operation, they will be activated. They receive signals as a result of the discrepancy between the sensory input and the efference copy that, together with the original potentiating signal, are sufficient to excite the relevant neuron beyond threshold. The muscle fibers in my little finger, however, may simply be unnecessary, so they do not receive further excitation beyond the potentiating signal; in fact, they may even receive depotentiating (inhibitory) signals from within the servomechanism. And though they are potentiated by neurons at a higher level in the hierarchy (because under some possible circumstances they could be of use in this project), they will not, under the present circumstances, be activated, because they do not receive excitatory signals from within the servomechanism.

Servosystems in general play a prominent role in robotics research. Some researchers in robotics (for example Brooks, 1987) claim that the use of servosystems can result in intelligent behavior without any need for appeal to representations and processes defined over them. Since I claim to be working within a fully intentional belief-desire psychology, it should be noted that I in no way endorse such a conclusion. First, Brooks does not intend for his robots to mimic the processes according to which human behavior is produced; and he is not concerned in the least with questions concerning behavior that is *in the first instance* autonomous or voluntary. Second, behavior at that simplistic level need not be bothered by competing high level interests; that is where processes defined over representations become important. Finally, and most importantly, Brooks, like others in AI and the cognitive sciences, appear to have no theory of representation, or if they do, it is very much like the causal or information theoretic accounts to which I

appealed earlier (see, for example, Pylyshyn, 1984, chapter 2). On such accounts there is no sense in which Brooks' robots operate *without* representations. They require perceptual inputs that clearly give rise to representational states on any plausible account of representation, and that are supposed to be matched up with efference copies. In other words, one cannot claim to have no use of representations by denying the representational character of the efference copies in servomechanisms.

In Gallistel's view, servosystems constitute the origins of intelligence in navigation (Gallistel, 1980, p. 10). The kind of framework called for here is one in which servosystems are nested within larger servosystems. Each servosystem is anchored by a node that receives potentiation from higher levels in the hierarchy. This is the neural structure that contains the efference copy. It also receives feedback from a sensory mechanism, which it passes on to lower nodes within the servosystem until the feedback it receives matches the efference copy. When this match is secured, it no longer potentiates lower nodes within the servosystem; its task has been completed. This servosystem may be just one of many nested within a larger servosystem that has its own efference copy, and which will continue to potentiate patterns of behavior until its sensory input matches its efference copy. Imagine an entire network of such structures and the plasticity of which it is capable. I share with Gallistel the suspicion that this type of organization can accommodate all of the voluntary behavior exhibited by even the most sophisticated organisms.[4] In the next section I will show how very little alteration of this framework is needed to capture the plasticity of intentional behavior.

Thus far we have seen how a single command can, depending on the prevailing circumstances, be executed in many different ways, each of which will have the same consequence. A circuit at a higher level in the neural hierarchy determines only a pattern of potentiation at the level beneath it. Which of these potentiated circuits is in fact activated depends largely on servomechanistic input. The proliferation of this patterning leaves for an enormous variety of realizations of the same command. I want now to show how this theory of motor systems can be extended to deal with intentional behavior, behavior explained by appeal to beliefs and desires.

The Physiology of Desire

Gallistel's treatment of the nature and origin of commands for intentional behavior falls under the rubric of motivational states. As he puts it,

[4] Gallistel, of course, discusses oscillators and reflexes in addition to servomechanisms. These other so-called behavioral units are, however, tangential to the concerns of this paper.

> Motivation . . . refers to those processes in the central nervous system that organize behavior so that, in the aggregate, the animal's separate acts tend toward some culminating point, or action, or state of affairs. Motivation refers, in other words, to those processes that impart to behavior the characteristics that make us speak of purposes and goals. (Gallistel, 1980, p. 321)

This characterization is somewhat cryptic. Gallistel is not sufficiently explicit about the origins of these commands, nor how they acquire the instruction-giving capacity with which he is more obviously concerned. It is my suspicion that we must look to the causes of these motivational states in order to determine the nature of the commands they issue, and why it is those commands that they issue.

I will give, from an evolutionary perspective, a preliminary account of the origins of these commands in terms of biologically relevant behaviors, and suggest later how this theory might be extended to biologically irrelevant goal-directed behavior. The initial search for the origins of these motivational signals, then, will be guided by the biological requirements of the organism. There can, in this context, be only one type of source for the motivational signals of which Gallistel speaks: a biological depletion, deficit or other state of imbalance in the internal environment of the organism. These signals may be, and in fact often are, caused indirectly by external objects like food or water. It is presumed that the motivational states owe their existence in part to natural selection. In most cases, excluding those where the mechanism involved is doing more than what it has been selected to do, there must be some biological merit to the motivational states. Let them be hormonally induced, if you will; it does not, for our purposes, matter. I will take it, then, that these motivational states, states that are causally implicated in the production of the potentiation patterns of coordinated behavior, are caused by the likes of, say, water depletion and low bodily temperature. It is signals arising from, for example, sensory mechanisms that detect extracellular fluid quality and quantity that give rise to what Gallistel calls the motivational states of the organism.

If we now take the hierarchy of servosystems of individual component motor systems and generalize to include other component systems, we will have achieved a level of plasticity significantly greater than that of which each individual system is by itself capable. Error detection in this larger system will not, of course, work exactly like that in a lower level servomechanism. Instead of an adjustment in limb trajectory, the result of error detection will be the continuation of potentiation of lower nodes in the hierarchy until no error is detected. We can only speculate as to what will in fact play the role of these higher level commands and efference copies. The present point is only that we need not employ any further organizational tricks to capture the increased plasticity of intentional (though still only biologi-

cally relevant) behavior. We are merely noticing what computer scientists have long recognized as the power of hierarchies.

A concern arises, however, if we adopt this account of the origin of motivational states. Suppose that sensors projecting to the hypothalamus register extracellular fluid depletion and issue signals as a result. The account that I am advocating would require this signal to impinge on some structure that is properly in the neural hierarchy that controls behavioral output. This is what I would take Gallistel to mean by motivational state, or, as he sometimes puts it, central motive state; a state that releases a potentiating signal when activated by a stimulus (Gallistel, 1980, p. 324). But if this is so, then, by some Ockhamesque principle, one might be driven to doubt the ontological status of such a structure. That is, it would seem unnecessary for there to be any identifiably distinct structure mediating the signal from the fluid detectors to the control hierarchy. To speak of motivational states as arising at higher levels in the control hierarchy is really to speak of them as arising from various quarters of the body, as they experience biological deficit. The potentiation signals are really nothing more than the proprioceptive analogues to sensory stimulation. Signals arising from the fluid detectors potentiate patterns in the hierarchy that will produce behavior that ideally will extinguish the signal.

This consideration, however, suggests a way to support the claim that there is a central motive state from which all potentiating signals must emanate (which is not to say, of course, that they must originate there). Quite likely, there are many signals issued as a result of biological deficit at any given time. Not every signal, then, can lead to behavior directed at the removal of the deficit. Some of these signals must be intercepted, while others are allowed to trigger behavior appropriate to the removal of the deficit from which they issue. I think it is plausible to identify a central motivational state with a circuit designed to perform such a function. There must be some interfacing of signals from the various quarters of the body such that the patterns of potentiation and depotentiation result in coherent behavior capable of reaching a single goal or set of consistent goals, rather than chaotic flailings robbed of any chance for success.[5] Something, it would appear, must receive these afferent signals and issue command signals; it is this type of circuit that I will call the central motive state (CMS).

As I will understand it, the CMS determines which instructions will guide the organism's behavior. It determines whether the organism is to seek food, water, shelter or some other commodity. It is, of course, crucial to the performance of such a task that it be sensitive to the input signals from the various proprioceptive sensors scattered about the body. And it is these input signals

[5] Essentially the same argument is given in Wise and Strick (1984, p. 446).

that I want to appeal to in characterizing desires naturalistically. I will view desires as neurophysiologically realizable functional units, conative states if you will, that take afferent signals from the proprioceptive sensors as inputs, and issue potentiating signals as outputs. It is the set of these functional circuits that I will call the CMS, where the CMS is responsible for the production of outputs that determine the ultimate consequence of intentional behavior.

The main argument for the claim that there must be a class of functional units mediating the signal from the proprioceptive sensor to the control hierarchy is that there must be something in virtue of which the potentiating signals issued from the CMS result in the patterns of potentiation that they do, patterns that eventually lead to the removal of the imbalance to which the sensors are responding.[6] There must, then, be some way for the fact that the signal arose from this or that sensor to play a role in determining where the potentiating signals exiting the CMS will flow, what patterns they will potentiate.

This functional unit must have a representational component that allows it to play the causal role that it does. What I am imagining is a neural structure that responds to a certain biological deficit. If the story I am about to tell is right, then it is a response to this deficit that explains why the structure in question causes what it does. Nature has selected such a structure in the same manner it has selected hearts. And if it is plausible to attribute functions to hearts, then too it ought to be plausible to attribute a function to this type of unit. This guarantees that there is some manner in which this type of unit is functioning "as it should."

Let us look at this more closely. The biological purpose of the signals issued by the proprioceptive sensors is to assist the organism in reaching some state that is conducive to the continued existence of the organism, a state, say, of sufficient fluid quality and quantity. It must, then, play some causal role in the production of behavior guided toward this end. Signals exiting the CMS must be directed to those circuits high in the control hierarchy that will potentiate patterns of behavior appropriate for, say, water acquisition. Undoubtedly, it is probably almost the entire range of an organism's behavior that is appropriate for food gathering, especially in higher organisms with complex food gathering strategies. In these cases, though the ultimate consequence of this behavior must be determined by the source of the proprioceptive signals, that whole range of behavior is potentiated, and, as we shall see, cognitive structures determine largely what particular behavior will be executed. For the sake of expository ease, however, I shall speak as if there are discreet water acquiring behavioral units.

---

[6] Empirical considerations that point to exactly this conclusion can be found in Stellar and Stellar (1985, chapter 4).

Since all of the many afferent signals from proprioceptive sensors must be channeled through the CMS, the signal from the water detectors cannot be given direct, unobstructable access to the high level water acquiring circuits. Its access to these water acquiring circuits must be given by a *proxy* in the CMS. That is, if there were only one proprioceptive signal, it could be hard-wired to the control circuits so that every time it was issued, it would potentiate an appropriate pattern. But since there are many signals vying for governance of the control hierarchy, this hard-wiring cannot be effected. Each signal must be such that it can be channeled through a single circuit (the CMS) and still potentiate the appropriate pattern. This can only be accomplished, I think, if there is a functional unit in the CMS responding to sets of proprioceptive sensors, those, say, that detect water quality and quantity. I envisage the relationship between these functional units to be such that only one (or a few) of them may perform its task at any given time; the others, as it were, are depotentiated. This is, of course, an oversimplification. There would have to be some way for structures to combine in such a manner to satisfy more than one desire at any given time. Perhaps those circuits that potentiate coherent patterns of behavior can be simultaneously active. In fact, for a suggestion of this very sort see Stellar (1985, p. 378). Each unit, then, is responsible for causing potentiation of a pattern appropriate to produce behavior that will acquire water. Thus, if the signal from the water detectors is to cause behavior that will replenish the organism, there must be some unit that responds selectively to the extent of the depletion in question, and that issues appropriate potentiation patterns when it is itself "potentiated."

A question outstanding is, of course, what determines, at a given time, which of the functional circuits can operate, which of them can perform its function. I think a quite plausible suggestion is that the answer be in terms of the relative strength of the afferent signals, where strength is coded perhaps in terms of the frequency of firing, or some other candidate code. Perhaps there is some mechanism that monitors the frequencies of the afferent impulses and potentiates the unit receiving the input signal of the highest frequency. There is obvious room for selective pressure in favor of such a mechanism since the strongest signal will be emitted by that proprioceptive sensors responding to the greatest deficit. The scheme I am suggesting would require, if the system is functioning properly, that the greatest need have priority of satisfaction; and this is sure to be a winning biological policy. This also gives us a way out of the circle of defining the strongest desire as that desire that causes action. The strongest desire is instead that desire that causes action when the conative faculties of the organism are functioning ideally, that is, as they were selected to function.

To extend this account to biologically irrelevant behavior we need only substitute for a unit responding to biological need some unit responding to a neural state correlated with a state of agitation or the absence of pleasure or satisfaction. Such units may have arisen through any number of evolutionary means: intensification of function and pleotrophy are two such examples of how biologically irrelevant conative states could have come to characterize our neural machinery. Also, I do not rule out the possibility that there is some process within the lifetime of an organism that gives rise to conative states of various kinds. Imagine happening upon a sports car and entering into a state we would ordinarily describe as the desire for this car. On the physiological account I am presenting, the perception of the car would have induced one into a neural condition corresponding to something like the absence of pleasure. A unit in the CMS responding to this condition then behaves in just the way that our biological units behave; the story remains the same from this point on. I cannot, of course, claim that this is in fact what goes on. My aim is merely to point out that no radical new theoretical machinery need be introduced in order to capture desires with no obvious or apparent biological relevance. The framework I sketch is intended to be general, the details of which may be other than I describe. I am not providing a full explanation of behavior, but a partial framework in terms of which full-blooded explanations may be cast.

We have before us, then, a picture of desires according to which they are proxies for various sorts of imbalances that sit atop a servomechanistically organized control hierarchy. We can see why it is that when we go to take out the garbage the voluntary movements that result in our lifting the garbage can off the floor do not constitute a consummatory piece of behavior, why we don't stop after having done just that much. Quite clearly it is because the object of the desire has not yet been reached. The conative state continues to issue potentiating signals to lower nodes in the hierarchy because it is still receiving excitatory signals from the imbalance to which it is responding. Consider how this works. The conative state is itself the beginning of a large servosystem of sorts within which smaller servosystems are nested. The conative state comes equipped with a component whose function it is to respond to a particular aspect of the internal environment of the organism. When that aspect is in a state of imbalance, the conative state will, provided nothing external interferes, issue potentiating signals until the sensory inputs from this aspect of the internal environment match a certain representation of it as being in balance, i.e., an error is no longer detected. We do not, therefore, stop moving after we have picked up the garbage can because the conative state that issued the signals that resulted in that activity are still registering a discrepancy between the sensory input from the neural correlate to, say, a state of dissatisfaction with the aroma of the garbage

and (what else should we call it?) the *efference copy*. If all goes well, the behavior will not stop until the sensory signals match this efference copy. This feedback loop works very much like our servosystem on a grand scale.[7] And this shows why it is that we continue to act until we have reached the object of desire. Desires are, after all, for what *will* satisfy us, not what we *think* will satisfy us.

Let me close with a speculation that might enhance the plausibility of the reduction I am proposing. I have maintained that the object of desire is determined by the function of a conative state, where that function specifies what the conative state causes under ideal or normal conditions. To supplement this approach we might look more closely at the representational component of the conative state, i.e., the efference copy. The efference copy is that component of the error detector against which the incoming signal is matched; matches serve to extinguish the potentiation issued from the conative state. On that ground, it is plausible to suppose that the efference copy is a representation of the object of desire; it represents what condition will satisfy the desire. What the efference copy represents might plausibly be said to be such external conditions as food, car-ownership and the like, rather than some state of internal imbalance. It will remain true, of course, that the units in the CMS will receive their signals directly from these states of internal imbalance, but, since external objects are correlated with these internal states, it is these that are represented by the efference copies; we may view this as nature's way of ensuring that we are impelled to act in ways that secure what we want. This interpretation of the representational content of the efference copy seems to complement the teleological functionalism discussed above.

Summary

My aim has been to demonstrate, by means of an example, that a belief-desire psychology is likely to survive the advance of neuroscience without being autonomous. Admittedly, only half the story (at best) has been told. But the physiological model of desire seems to preserve and explain much of what was revealed to us in our examination of the role of desires in belief-desire explanations of voluntary behavior. We may not *know*, as Hubel suspects, what it is we are talking about when we probe the physiology of mind. But if the thesis of this paper is right nevertheless, we just might know what it is we are talking about when we appeal to beliefs and desires to explain our behavior.

[7] See Stellar and Stellar (1985, chapter 4) for a discussion of the neural structures involved in such a framework.

References

Brooks, R. (1987). *Intelligence without representation*. Unpublished Manuscript, Massachusetts Institute of Technology.

Churchland, P.M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*, 67–90.

Churchland, P.S. (1986). *Neurophilosophy*. Cambridge, Massachusetts: MIT Press.

Clark, A. (1980). *Psychological models and neural mechanisms*. Oxford: Clarendon Press.

Dennett, D. (1987). Evolution, error and intentionality. In D. Dennett (Ed.), *The intentional stance* (pp. 287–321). Cambridge, Massachusetts: MIT Press.

Dretske, F. (1988). *Explaining behavior*. Cambridge, Massachusetts: MIT Press.

Fodor, J. (1976). Special sciences. *Synthese, 28*, 77–115.

Fodor, J. (1987). *Psychosemantics*. Cambridge, Massachusetts: MIT Press.

Forster, M. (1987). *In defense of a causal theory of representation*. Unpublished Manuscript, University of Wisconsin–Madison.

Gallistel, C. (1980). *The organization of action*. Hillsdale, New Jersey: Erlbaum.

Gallistel, C. (1981). Précis of *The organization of action*. *Behavioral and Brain Sciences, 4*, 609–650.

Ghez, C. (1985). Voluntary movement. In E.R. Kandel and J.H. Schwartz (Eds.), *Principles of neural science* (pp. 487–500). New York: Elsevier.

Ghez, C., and Fahn, S. (1985). The cerebellum. In E.R. Kandel and J.H. Schwartz (Eds.), *Principles of neural science* (pp. 502–521) New York: Elsevier.

Helmholz, H. von (1962). *Treatise on physiological optics, Volume III*. New York: Dover.

Hooker, C. (1981). Towards a general theory of reduction, Part I. *Dialogue*, March, 33–59.

Hubel, D. (1979). The brain. *Scientific American*, September, 39–47.

Jackson, F., and Pettit, P. (1990). In defense of folk psychology. *Philosophical Studies, 59*, 31–54.

Kim, J. (1989). The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association, 63*, 31–47.

Matthen, M. (1988). Biological functions and perceptual content. *Journal of Philosophy, 85*, 5–27.

Merton, P. (1972). How do we control the contractions of our muscles? *Scientific American*, May, 30–37.

Millikan, R. (1984). *Language, thought and other biological categories*. Cambridge, Massachusetts: MIT Press.

Papineau, D. (1984). Representation and explanation. *Philosophy of Science, 51*, 550–572.

Pylyshyn, Z. (1984). *Comptation and cognition*. Cambridge, Massachusetts: MIT Press.

Russell, J.M. (1984). Desires don't cause actions. *Journal of Mind and Behavior, 5*, 1–10.

Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.

Stampe, D. (1977). Toward a causal theory of linguistic representation. In P. French, T. Euhling and H. Wettstein (Eds.), *Midwest studies in philosophy* (pp. 81–102). Minneapolis: University of Minnesota Press.

Stampe, D. (1986). Defining desire. In Marks, J. (Ed.), *The ways of desire* (pp. 144–164). Chicago: Precedent.

Stampe, D. (1988). Need. *Australasian Journal of Philosophy, 66*, 129–160.

Stellar, E. (1985). Brain mechanisms in hedonic processes. In D. Pfaff (Ed.), *The physiological mechanisms of motivation* (pp. 465–479). New York: Springer-Verlag.

Stellar, J., and Stellar, E. (1985). *The neurobiology of motivation and reward*. New York: Springer-Verlag.

Stich, S. (1983). *From folk psychology to cognitive science*. Cambridge, Massachusetts: MIT Press.

Wise, S., and Strick, P. (1984). Anatomical and physiological organization of the non-primary motor cortex. *Trends in Neuroscience, 7*, 442–446.

Wright, L. (1973). Functions. *The Philosophical Review, 82*, 139–168.