ABSTRACT. In this paper I sketch an account of (a) the origin of the terms and concepts of folk psychology, and (b) the true nature of mental states. I argue that folk psychology is built on metaphors for the functioning physical body, and that mental states are neurological traces which serve as schematic 'mental images' of those same functions. Special attention is paid to the folk psychology of self-consciousness. In particular, I argue that the notion of introspection is mistaken, and I criticize recent claims of Patricia and Paul Churchland on this subject. I conclude by discussing recent empirical evidence in support of my approach.

Eliminative materialism claims that folk psychology is a false theory and that the elements in its conceptual framework—episodes of believing and desiring, intentions, pains, self-consciousness, etc.—are not reducible to elements in neuroscience because they do not pick out real natural kinds, or are not well-formulated theoretical elements. Paul Churchland likens these concepts to those of witches or phlogiston: with advances in knowledge these were seen to refer to nothing. This fate may or may not lie in store for the concepts of folk psychology. But the very idea that we could decide that there is no believing or desiring in the same way that we decided that there were no witches arouses resistance because of the universality of these sorts of concepts, the tradition behind them, and their apparent usefulness. What sort of change could induce us to agree that we are wrong about them? What could replace them?

This resistance applies most strongly to two properties accompanying many mental states postulated by folk psychology: intentionality and consciousness (in the sense of self-awareness or introspection). If eliminative materialism is true then we are wrong in thinking that these properties are irreducible real essences of some mental states. In this paper I suggest specifically how we might be wrong. I propose that all mental terms and concepts are metaphors for the functions of our physical bodies, and that mental events actually consist of detailed or schematic neural traces of these same functions. The metaphors lead to philosophical perplexities when taken literally.

If this proposal is right the consequence is a blend of reductionism and eliminative materialism. Some mental concepts, including those involving intentionality, may be reducible to physiological concepts; oth-
ers, including that of a sensation about which there can be perception-like knowledge, may refer to nothing. In what follows I make substantive use of mentalistic idioms such as "think" or "imagine;" however, I do so only in contexts in which I argue that these idioms refer to events reducible to types of brain events. Terms such as "theory" or "metaphor" are also mentalistic, in that holding theories or using metaphors requires intentionality. If my analysis of intentionality is right, then such notions are also, in the long run, reducible or eliminable. But it is necessary to use them for the time being, or the ideas in this paper could not be coherently expressed.

Regarding mind-brain reductionism: to save time I will start by assuming a questionable claim instead of arguing for it. This is that the reducibility of neuroscience to chemistry and physics is not so problematic as the reducibility of psychology to neuroscience. It could be that multiple realizability arguments could be made against the reduction of neuroscience to chemistry and physics—other creatures could have neurons without having our chemistry—but I am going to ignore that possibility. I feel safe in doing so because current nonreductionist psychologies which I oppose argue almost exclusively against the latter.3

I make this assumption because I will argue that certain psychological elements in the mental processes postulated by folk psychology can be reduced to biological processes, common to living organisms, and that these processes do not confront us with a mind-body problem.

I will begin by looking at some similarities between mental and physical concepts which I claim provide a clue to the nature of mind. I then explain my approach and show how it handles both intentionality and consciousness. In connection with the latter I discuss a difficulty with a reductionist account of qualia recently proposed by Patricia and Paul Churchland.

I. THINKING IS LIKE ACTING ON SEEN OBJECTS

Dennett has argued that all intentional idioms posit something like a little man in one's head, interpreting input, knowing where to find information, sending signals, etc. He argues convincingly that this "homunculus" must be "exorcised" before one can have a complete explanation of mental acts. Dennett's point is that everyday mental idioms presuppose or "implicitly posit" the intelligence they supposedly explain, and not that we really talk about a little man in our heads.4 But in fact our ordinary concept of the mental is of a realm in which we perform actions parallel to all those that we perform bodily, on similar sorts of objects. This will be clear if we look at various parallel mental and physical concepts.

First, there are both mental and physical "acts". Most mental acts, like some physical ones, essentially involve objects. Physical acts of approaching or avoiding objects, manipulating them, making and altering them, or just focusing one's eyes on them, have counterparts in the mental acts of concentrating on or repressing thoughts of objects, analyzing them, creating or developing them, or just thinking about or imagining them. Mental acts can be occurring or dispositional, as can physical acts. I can take a mental, as well as a physical, 'attitude' or 'stance' with respect to something.
Second, like physical acts, mental acts have a goal or focus, and are often said to be directed at "intentional objects". The 'object' in this sense is, for example, the idea I am trying to recapture or the fountain of youth I am looking for. In these cases, the being of the object is to be thought about. Other objects of mental acts can have independent existence: I can 'look at a book or taste an apple.

Third, there is said to be mental, as well as physical perception. The most common mental metaphors describe consciousness and knowledge as seeing. Mentally as well as visually we can 'focus' only on one thing at a time: distractions interfere with one's conscious deliberation about a problem. Also common are mental 'feelings' and 'sensings'.

Finally, mental and physical objects are given values, and these values motivate action. Propositions are true or false; arguments valid or invalid; ethical concepts good or evil. These values motivate or are offered as justification for the respective mental acts (or propositional attitudes) of affirmation or denial, acceptance or rejection, approval or disapproval.

Thus, as they have traditionally been described, mental events parallel bodily ones. Mentally we perform the same abstract action patterns and enter into the same types of relations to objects as we do physically. We have mental goals, suffer mental pain (sorrow) and pleasure (joy), experience chronic mental conditions (moods), have special mental abilities (intelligence or talent), and can be mentally tired (depressed) or energetic (manic).

Before beginning to make something of these parallels I must consider two important objections. First, why take these metaphors seriously? Maybe all they show is that we don't know the real nature of psychological processes and entities, and so fall back on terminology familiar from another context. This might be the response of eliminative materialism.

Here is the problem with this response. A great many of our theories and concepts concerning the mind lend themselves to physical language. Now either all mental concepts can be expressed this way, or some, namely concepts of intentionality and consciousness, cannot. If all can, then by pursuing the metaphor we should be able to elucidate apparently mysterious mental events in terms of familiar physical ones. If some cannot, then there is a problem about the origin of these concepts. Where are they supposed to have come from?

Some functionalists claim that intentionality arises within systems of a certain degree of complexity and that it is not reducible to nonintentional ideas; our ideas of intentionality come from experience of it. Functionalism has had little success with consciousness, but a functional account has not been ruled out.

For eliminative materialism, to say that some mental concepts are not reducible to physical ones is a more serious step. Even if what passes for intentional activity or consciousness could be shown to be complete unlike the traditional accounts the problem remains: how did we get the idea in the first place? Because this question must be answered in a complete theory of mind, I think it would pay to take the meta-
phorical nature of mentalistic idioms seriously and see how far it can be pushed.

The second objection attacks my comparison of mental and physical events. It will be objected that in my comparison of mental and physical action, perception, etc., I have not distinguished between the mental and the nonmental, since the physical actions described are all intelligent, goal-directed actions. The real problem is not the difference between thought and action but the difference between intelligent and non-intelligent systems, and there are no significant parallels between these.

Admittedly, the physical actions I have described are goal-directed. But as I indicated earlier, explaining goal-directedness per se is not the real problem in philosophy of mind. As Dennett implies, there is a continuum in biology from simple to highly complex well-designed, goal-directed systems, from cells to mammals. In general the problem for such systems is the problem of how biology or organic chemistry is reducible to inorganic chemistry and physics, not the problem of mind and body. As Dennett says, a system is mental or intentional "only in relation to the strategies of someone who is trying to explain and predict its behaviors"; this account applies to attempts at self-understanding by the system in question as well as to attempts at understanding others. This paper is about mind and ascriptions of mentality, both of which are created when humans ask self-conscious questions about themselves and each other.

The essential differences between the mental and the physical acts described above is that the physical acts (a) are all directed at real, existent objects, and (b) do not presuppose self-awareness but only object discrimination. The mental acts, on the other hand, (a) do not require independently existing physical objects, and (b) sometimes entail self-awareness. In other words, only in the above mental acts are found the key aspects that have traditionally resisted full explication by all theories: intentionality (in Brentano's sense) and self-consciousness or conscious awareness of qualitative states.

Thus the objection that demands the most careful response is that while thinking may be like acting in some superficial way, perhaps because of an accident of language, it is essentially different. I therefore have to show that both intentionality and self-consciousness can be analyzed in terms of processes or states of our physical bodies.

II. INTENTIONALITY

According to Brentano, mental processes are characterized by "direction upon an object", even though the object need not exist. If a thinker thinks about something, only the thinker need exist. Relations between physical things, what Brentano calls "strict" relations, require that both relata exist. Thus mental relations are not reducible to physical relations. Brentano's distinction between mental and strict relations underlies many current nonreductionist theories.

Let us consider an example. Suppose that yesterday I was bitten by a dog and today I am thinking about the dog that bit me. Brentano's analysis might be as follows: yesterday there were two relata, I and the
dog, both of which had to exist for the relation "bit" to hold. Today only one of the relata, I, must exist for the relation "thinking about" to hold. Thus thinking about is (irreducibly) mental and biting is strict.

There is another analysis that applies equally well to at least one way I might be thinking about the dog. Yesterday I was bitten by it, today I imagine or remember being bitten by it, through a mechanism which activates neural traces of response patterns encoded by the event yesterday. A new stimulus signals similar response patterns through a network of associations, in a process which functions automatically unless it is inhibited or eclipsed by competing stimuli. If the traces are reactivated exactly as they were encoded, we say that I remember the event; if there are alterations or substitutions, we say that I merely imagine it. Either way I am thinking about the dog. But at least in this instance of intentionality Brentano’s asymmetry between mental and physical events is avoided. Yesterday there was a real physical relation between me and a real dog; today there is a remembered physical relation between me and a remembered dog. Instead of construing the act of remembering widely, as a relation between the subject and an external object, we can construe it narrowly as a brain event which recapitulates an earlier physical act; act and object are both represented in the brain activity.

Intentional acts concern future as well as past events. If imagining is reactivating traces of what happened in the past, how can we explain future-directed intentional acts? We can do this by supplementing the account of imagining with an account of the response imagery involved in volitional action. Recent work on brain mechanisms of movement has supported William James’ "ideo-motor" theory. James claimed that:

> When a particular movement, having once occurred in a random, reflex, or involuntary way, has left an image of itself in the memory, then the movement can be desired again, proposed as an end, and deliberately willed. ... We may consequently set it down as certain that, whether or not there be anything else in the mind at the moment when we consciously will a certain act, a mental conception, made up of memory images of these sensations, divining which special act it is, must be there.⁹

James theorized that these motor images will function to bring about the imaged action unless the action is deliberately inhibited. On this account, thinking about a future action would involve activating the appropriate motor imagery and at the same time inhibiting the action. Recent work by Edward Evarts and others¹⁰ supports this theory with the discovery of abstract action patterns in the motor cortex, which function to initiate voluntary action. The feeling of volition accompanying the mental act of planning for the future is the occurrence of this action imagery. This work will be discussed in more detail in Section IV of this paper.

It can be objected that my account treats intentional states only insofar as they have nonpropositional objects. Dennett argues that Brentano’s analysis can and should be reformulated entirely in terms of propositional attitudes. He wants to analyze all intentional states into propositional objects because this analysis will best account for the lack
of independent existence of these objects. Only a proposition is "a different thing under different descriptions". The objection is that if Dennett is right my example fails because the 'object' of my imagining, if it is simply reactivated neural traces of the previous event, is not propositional; it has no semantics. Only if I image that I was bitten by a dog can the process be analyzed propositionally. But if imagining, in my example, is imagining that, then it is intentional in a purely mental sense, since it is more than just reactivating memory traces. Thus according to this view only if I analyze imagining as mechanical will my example work, but then I have not analyzed an intentional process.

This objection brings out a key point in my analysis: when I think about a dog by imagining being bitten by it, the dog as intentional object is dependent on the mode of representation in the same way it would be dependent on a description. Imagining the dog, in my example, is recapitulating my original response to the dog. But my original response was not to every aspect of the dog, but just to certain salient features: its teeth and not its markings; its laid-back ears and not its white feet. These were what I focussed my eyes on and what triggered the adrenalin flow. Now, the reactivated memory traces include not only what was present in the visual field (I can remember what the dog's spots look like if I try) but also my entire set of responses, which pick out one aspect of the event from other possibilities. Thus I am imagining the dog as menacing or dangerous, and not, say, as attractively marked, although the visual information necessary for that interpretation is available.

The ability to pick out one aspect from others in a set of stimuli by selective responsiveness can be given a biological analysis. When aplysia recoils from a probing needle or a neural crest cell migrates to its destination in the developing embryo certain aspects of the environment are salient as triggers of (temporarily) dominant mechanisms; other stimuli, which in other circumstances could play analogous roles, are in this state of the organism or cell 'ignored'. Imagining an object under a description, or taking an attitude toward a propositional object, can be analyzed as the same sort of goal-oriented process as we find in these simple life forms. We 'notice' certain features of our environment as responding selectively to these features, and we 'ignore' the rest of the environment by not responding in that way to it. We can then recapitulate these responses. Thus worries about the ontological status of intentional objects which are dependent on descriptions can be allayed with the recognition that these 'objects' are patterns of neural responses and do not require a special propositional analysis.

On our ordinary notion of intentionality, states like believing or imagining are commonly identified in part through external relations. What it is that I am imagining is determined by the context in which the imagining occurs; in other words, imagining is construed widely. Stephen Stich argues that the folk concept of belief would individuate beliefs differently in a case where two people had identical inner states but the objects of the beliefs were nonidentical. The same can be said for imagining. For example, suppose you ask me to imagine being bitten by a dog we see in the distance, which looks like one that bit me several years ago. I might imagine that this current dog is biting me by reactivating traces of the previous event, and I might do so without realizing that my current imagery is a memory of the previous event. In this case we would say that I am imagining this dog, and not the previous one,
even though the inner events in the two cases are indistinguishable. In other words, identical mental state tokens can belong to different mental state types depending on external relations such as reference and context.

If our ordinary intentional notions are wide in this sense, how can they be captured by a 'narrow' analysis? They can't, of course. On this matter I agree with Fodor, an advocate of a narrow construal, who says "truth, reference and the rest of the semantic notions aren't psychological categories. What they are is: they're modes of Dasein". As such they are the proper subject of a discipline other than psychology.

The fact that a dog I see reminds me of another one that bit me is not, of course, without psychological interest. But a fact can be relevant to psychology without being a purely psychological fact. In suggesting that we do not have to worry about common sense, wide notions of intentionality in psychology, I am not giving up the claim that what has seemed essential and irreducible about these notions must be explained. Common sense notions may be, as Stich puts it, 'hybrids', containing both narrow (psychological) and wide (legal, historical, geographical, etc.) components. My purpose in this paper is to isolate the components belonging to psychology, and suggest how they are rooted in biology. In the case of intentionality, the component is goal-directedness, which I have explained in terms of selective responsiveness. It is this element which has seemed uniquely human, and which has led to such excesses as emergentism. Thus I differ from Stich and other eliminativists who make no serious effort to explain the intuitions which lead to folk psychology. The intuitions are clues to be treated with respect. Thinking seems like acting, and is so described in folk psychology metaphors, because the two are homologous. Thinking cannot be understood without recognition of these roots.

My examples of thinking about a dog by imagining a single event is extremely simple, and it can be objected that in cases of thinking that involve problem solving, evaluation, etc., a biological analysis would be inadequate. My rather radical claim is that all mental states and processes involving intentional objects can be analyzed in the same way: they are multi-level recapitulations of the neurological components of earlier physical acts and responses. If these recapitulations include all biological reactions like eye focus and patterns of stimulus-response relations, then under varying degrees of abstraction they can underly more complex or abstract cognitive processes. As we have seen, mental processes in general can be analyzed in terms of types of acts performed on types of objects. Thus a necessary condition for greater complexity or abstraction would be the ability to form schematic or prototypical images of acts and responses. Empirical evidence of this ability will be discussed in Section IV.

II. CONSCIOUSNESS AND QUALIA

Traditionally the most puzzling aspect of mind has been consciousness. This is a broad concept; it can refer at one extreme to any non-comatose condition involving mental activity, and at the other extreme to direct awareness of one's own concurrent mental states. David Armstrong has distinguished three states: minimal consciousness, where there is some mental activity such as a fleeting sensation; perceptual conscious-
ness, where there is awareness of what is going on in one's environment and one's body; and introspective consciousness, which is "perception-like awareness of current states and activities in our own mind". A central problem for materialism has been to account for the qualitative features of introspective consciousness, and many believe that this cannot be done in a physicalist framework. Thomas Nagel in "What Is It Like To Be A Bat?" argues that because conscious experience is always 'appréhended' only from a subjective point of view, an objective scientific approach to qualitative features of such experience will never capture all the facts.

I will introduce my account of consciousness with an examination of some recent alternative theories, beginning with Armstrong's explanation of "introspective consciousness". He cites the example of a long-distance truck-driver who 'comes to' after driving for a period without realizing what he was doing. Before coming to he had minimal and perceptual consciousness; afterwards he has introspective consciousness, a sort of inner perception of his mental states.

Armstrong's account postulates both mental states which are accessible as objects of awareness, analogous to objects of external perception, and a faculty of mental perception by which we can gain information about mental states. For the truck-driver, coming to is beginning to activate his inner perception. Armstrong emphasizes, along with other philosophers influenced by Hume's "bundle theory", that introspective consciousness does not present us with a self along with perceptions and other states that we are aware of. But in that case, how can we account for the special quality of introspective awareness as self-awareness?

Armstrong's answer is that introspection is "theory-laden", and that we learn to observe mental states and activities as properties of a unitary thing. The level of theory involved is not necessarily sophisticated; it is similar to the "theory" that a tomato has "sides and back, top and bottom, a certain history, certain causal powers, and these things certainly do not seem to be given in perception". He implies that this theory has a phylogenetic, not an ontogenetic, source. Our ability to introspect facilitates integration:

If we have a faculty that can make us aware of current mental states and activities, then it will be much easier to achieve integration of the states and activities; to get them working together in the complex and sophisticated ways necessary to achieve complex and sophisticated ends.

Given this function of introspection, the idea of the unity of the self follows:

We can now understand why introspection so naturally gives rise to the notion of the self. If introspective consciousness is the instrument of mental integration, then it is natural that what is perceived by that consciousness should be assumed to be something unitary.

In its general outlines this account is plausible. The phenomenological descriptions of different levels of consciousness are convincing, as is the account of the evolutionary advantage of integration of states of the
self. However, Armstrong's insistence that we have a special faculty for perceiving our mental states is problematic. Apart from the empirical fact that there are no known mechanism of introspection which are different from the mechanisms of sense perception and proprioception, there is the difficulty pointed out by Sydney Shoemaker. Perception is mediated by sense impressions caused in us by perceived objects, and these impressions explain our knowledge of the objects. This means of acquiring knowledge leaves open the possibility of misidentifying the objects we perceive. But since there are no sense impressions of sense impressions, and no possibility of misidentifying the self of whose sense impressions we have knowledge in introspection, then introspection cannot be a form of perception. 22

Why would Armstrong think of introspection as a form of perception? Shoemaker's explanation is that Armstrong holds such a broad theory of perception that simply getting information from something can count as perceiving that thing. Shoemaker offers an alternative account of introspection which explains the "immunity to error through misidentification" of the self known through introspection, and also explains how introspection can yield self-knowledge. Introspection gives us "direct knowledge of facts about ourselves," 23 rather than knowledge of nonfactual objects.

These facts are not mediated by sensations through which we know them, and hence are not perceived. Rather

we are so constituted that our being in certain states directly produces in us beliefs about ourselves to the effect that we are in those states. 24

An important difference between Shoemaker and Armstrong is that, while for both mental states such as sensations are objects of knowledge, for Shoemaker there is no special phenomenology to self-awareness:

Now some of the states we are aware of in introspection are themselves sensory states having a phenomenal character. And in the case of these it is natural to say that there is "something it is like" to be aware of them. But it seems plain that this "something it is like" is just the phenomenal character of the states themselves, and not the phenomenal character of still other states that are sense impressions of them. 25

For example, if I am looking at a red object, and am also introspectively aware that I am having a sensation of red, the phenomenal quality of my introspective awareness is nothing more than the phenomenal quality of the sensation of red—that is, the quality produced in me by the red object—and not a further quality through which I gain introspective knowledge of the sensation of red. Shoemaker does not explain the mechanism of introspection. He uses the term to refer to a state in which one makes use of knowledge of one's mental states, in order to make further inferences, verbal reports, etc. Introspection for Shoemaker seems to be functionally rather than experientially characterized.

Shoemaker's criticisms of Armstrong's perceptual model are convincing if one accepts, as seems reasonable, the observation that there are no sense impressions of mental states. However, his rejection of
Armstrong's account leaves a gap: there is an experiential quality to introspective consciousness not present in perceptual or minimal consciousness, and the experience is perception-like, readily captured by perceptual metaphors such as Armstrong's:

Introspective consciousness seems like a light switched on, which illuminates utter darkness. It has seemed to many that with consciousness in this sense, a wholly new thing enters the universe.26

Shoemaker's account of introspective consciousness lacks an explanation of "what it is like" to be conscious. However, we do not want an explanation in terms of inner perception of our sensations by means of sense impressions of them. Is there a way to account for the qualitative feature of consciousness while avoiding the perceptual model?

I think we can get what we want by elaborating on the mechanism by which introspective consciousness is achieved. In spite of their differences, both Armstrong and Shoemaker agree that the mechanism is not that of ordinary sense perception. I want to argue that introspection is accomplished by just the mechanisms of sense perception and proprioception that we normally use to get information about the physical world, and that introspective consciousness consists in the reentry of traces of past perceptual experiences (by means of reactivated neural response patterns) into the perceptual system, to be processed along with new sensory information.

My account draws heavily on two recent, and superficially very different, theories of self-consciousness. One is that of Gareth Evans in *The Varieties of Reference*.27 He argues that self-awareness involves the reuse of perceptual mechanisms, along with perceptual imagery which provides a self-image and world model (or "cognitive map") by which we identify our current experiences as those of a person among others in the objective spatio-temporal world. Evans' account is, in its general approach, similar to Shoemaker's, but he leaves room for a phenomenology of self-awareness (although he does not elaborate on this aspect). The other is a model for a neurological mechanism for self-consciousness, in *The Mindful Brain* by Gerald M. Edelman.28 As he summarizes the model:

The basic idea is that the brain is a selective system that processes sensori-motor information through the temporally coordinated interactions of collections or repertoires of functionally equivalent units each consisting of a small group of neurons. According to the model developed here, the brain processes sensory signals and its own stored information upon this selective base in a phasic (cyclic) and reentrant manner that is capable of generating the necessary conditions for conscious states.29

The groups of neurons consist of 50-10,000 individual cells; there may be $10^6$ such groups in a particular region of the brain. The groups are characterized by "diverse intrinsic connectivities but similar extrinsic connectivities" which allow sets of groups that are distinct but functionally equivalent. There is a "primary repertoire of neuronal groups formed during ontogeny and early development", which functions in recognition of new input and of the activity of other neuronal groups; recognition consists of alterations in patterns of firing in response to such
stimuli. Selection by experience produces a "secondary repertoire" within the primary repertoire; this consists of subgroups whose response to new input has become more likely due to repetition in experience. Response patterns activated in recognition can themselves be recognized; in this way recognition events are "nested". Past recognition events are thus summed or abstracted in higher order events.

Activation of past recognition events in response to new input amounts to the phenomenon of associative memory. These associations, in the form of multimodal abstract patterns, interact with multimodal input to generate the conscious state. Specifically, the interaction is made possible by "reentrant signalling": the system "is designed so that an internally generated signal is reentered as if it were an external signal". In other words, when one is conscious, past associations become part of current perceptions.

Edelman points out that while the model may yield sufficient conditions for conscious awareness, the issue of the quality of sensory modalities may not be completely accounted for:

... it is likely that the sufficient conditions for awareness arise from a historical process in each individual whereby increasingly abstract routines are placed in the secondary repertoire. The additional possibility must be entertained, however, that early associations with those areas of the brain concerned with affective states are critical in distributing into storage a series of response patterns which are sampled later in life. Such patterns may lead to chemically mediated changes in a variety of somatic responses. Although their "quality" is not discussible in scientific terms because the only possible access to these responses is indirect (behavioral observation) or verbal (possibly at a time when verbal communication have not yet developed), later verbally reportable "qualities" may be tied to their responses by abstraction through "recognizing" groups.

I interpret Edelman to suggest that while much of conscious awareness can be accounted for in terms of response patterns with a relatively recent history, which could be traced by objective means, some abstract features of these responses are survivals of early affective associations; the history of these is inaccessible by science. It would seem, however, that it is only in practice, and not in principle, that these physical features of the brain cannot be studied. The model under discussion does not provide a way to trace them, but they are no different in kind from other response patterns underlying conscious experience.

This model can be adapted to provide an account of introspection which combines Armstrong's phenomenological insights with Shoemaker's recognition that we do not literally perceive our conscious states. I propose that introspective consciousness consists of the activation of "images" or response patterns from the memory store (the primary and secondary repertoires). These images represent past perceptual, proprioceptive, affective and motor states, comprising all sorts of experiences which give me information about my self: the appearance of the world from my position in space, the state of my body, and the actions I am performing. These sorts of images represent me "from the inside". As
Evans notes, an adequate self-concept requires a sense of myself as a person like other people in the objective world. Thus there must also be, associated with these subjective images, objective ones representing a person perceived "from the outside"; person "prototypes", which are activated along with subjective images when I am thinking of myself as a person among others in the objective spatio-temporal world. The simultaneous reentry of selected past experiences along with new stimuli creates the experience of introspective consciousness.

A striking feature of this experience is captured by Armstrong's description: it is "like a light switched on". I think that what this description points to is the fact that introspective consciousness is the processing of "layers" of images along with sensations; these give an extra dimensionality or depth, so to speak, to ordinary perceptual consciousness. The past perceptions reenter the system in the form of response patterns, along with the current input, and the effect of the combination is to create the illusion that perceptions themselves are being perceived.

One advantage of this model is that it allows for degrees of self-consciousness; 'how conscious' one is depends on how much imagery is associated with one's current perceptions. It accords better with the model and with actual experience to postulate a continuum of degrees of consciousness, rather than three or four discrete states, as Armstrong does. However, the model does explain the striking experience of "coming to" in the truck-driver example. Sudden arousal or the appearance of novelty can trigger associations which are not activated by habitual behavior in familiar settings. As Armstrong describes it, bare perceptual consciousness may operate without any activation of self-imagery, and hence without being incorporated into the basic store, which is updated when it is activated in association with new sensory input. If sensory input does not activate the store, it is not encoded into it. Thus, when he comes to, the truck-driver cannot retrieve memories of actions performed while he was only perceptually conscious. (He might still retrieve the memories in some unusual way, such as by direct electrical stimulation of portions of his cortex.)

Another advantage of the model is that it accounts for much of the mystique surrounding conscious experience. The main difficulty with the term "image" is that it is usually used for distinct and recognizable visual images. But images can be nonvisual, and they can be vague, schematic, sketchy, and hence unrecognizable. One may have a vague feeling of seeing or doing something, without being able to identify the feeling. This vagueness lends itself to the creation of myths or illusions about the nature of consciousness, such as that expressed by the perceptual model of introspection. Self-consciousness involves the activation of past sensori-motor experience; thus it includes images of motor responses such as eye focus and head turning associated with visual attention. The images, while indistinct, give the impression that something is being perceived. Thus we hypothesize an 'inner sense' which perceives mental states, analogous to the outer sense which perceives objects. In fact, what seems to be inner sense is just a vague image of outer sense.

This model captures the important features of Armstrong's theory, while replacing his postulation of a special faculty of inner sense with a mechanism requiring only outer sense and memory. This has the advan-
tage of parsimony, as well as that of avoiding the serious difficulties with the perceptual model. It also goes beyond Shoemaker's account in allowing for a phenomenology of self-consciousness. There is something "it is like" to be aware of one's mental states, and that is more than just the phenomenal character of the states themselves. This "something it is like" is not, as Shoemaker points out, the phenomenal character of impressions of these states. However, it is the phenomenal character of other, past states, given in imagery. This account unpacks the notion of our being "so constituted that our being in certain states directly produces in us beliefs to the effect that we are in those states". The process is direct in that it is noninferential and unmediated by sense impressions of the current states. But in another sense, the production of beliefs about one's current states is indirect. Something more is required than just a perception for there to be awareness that one is perceiving; there is self-awareness in addition to the perception. Self-awareness is not, of course, perception of a self analogous to the perception of an external object, but there is additional perception in a sense. Self-awareness is the activation of past perceptions, including bodily perceptions and motor feedback, from a store of such perceptions. When these images, summed and abstracted through recognition by various levels of neuronal events, reenter the system as input along with new perceptions, we can say that one is aware of one's self perceiving, and that this awareness is a different sort of state, involving quantitatively more perceptual activity, from a state in which one is simply perceiving an external object in perceptual consciousness. Thus there is a view according to which, in one sense, Armstrong and Shoemaker are both right about perception and consciousness, and in another, neither is.

A third advantage of this view is that it offers hope of an at least partially objective approach to the difficult issue of sensation qualia. At first glance, this advantage is not obvious. Understanding how past perceptual experiences are summed and re-experienced along with new input does not account for such qualitative properties as redness or the throbbing of a pain. The most troubling feature of such properties is their apparent unanalyzability. While it may be acceptable, even if only barely, to analyze one's feeling of self-unity in experience as the product of many different images, it seems impossible to attempt the same sort of analysis of the experience of red. This intuition of the un-analyzability of sensation qualia is expressed in the inverted spectrum problem frequently viewed as a serious difficulty for functionalism. On my view as well, it would seem that two people, a and b, could have identical states of self-awareness, composed of identical sorts of imagery superimposed on identical sensory input, except that the sensations of a are experienced as having the quality he has learned to call "red", while the sensations of b are experienced as having the quality a would call "blue". If b has learned to call the quale he experiences "red", then there is an important feature of subjective experience which has not been captured by my account.

The present approach cannot dissolve all the philosophical problems connected with qualia, but it offers a partial solution. The solution is based on the premise, suggested by the view that consciousness of our mental states consists in imagery activated in association with those states, that our conscious qualitative experiences are composite. What seem to be unitary and unanalyzable phenomena, such as the awareness of red or a burning sensation, in fact consists of multimodal sensori-mo-
tor and affective associations, both learned and, in cases such as aversive responses to pain, biologically predetermined. The quale is largely functionally definable, in terms of its causal links with these associations. If this is right, then if a and b have identical responses to functionally identical input, they have, to a large extent, qualitatively identical experiences.

There is evidence that some sensation qualia are composite in this way. For example, while introspection presents pain as unitary, its aversiveness inseparable from its particular phenomenal quality, advances in anesthesiology reveal that the aversiveness is a separate element. Dennett points out that in the case of analgesics like morphine or nitrous oxide after receiving the analgesic subjects commonly report not that the pain has disappeared or diminished (as with aspirin) but that the pain is as intense as ever though they no longer mind it.33

Furthermore, pain can apparently exist in both its phenomenal and aversive components when the subject is unaware of it. Some anesthetics allow pain stimuli with both these components to reach the neocortex, but depress the reticular formation (which arouses the subject, making possible introspective consciousness):

The shutting down of the reticular formation by anesthetics . . . prevents or depresses "recruitment" by those stimuli; they arrive at the cortex, but do not produce the normal spreading ripple of effects; they die out.34

What this suggests is that much of the felt qualitative nature of a single episode of pain may be combination of aversive components plus associations in the form of sensori-motor imagery which are activated during normal self-conscious awareness of pain. This means that the actual unanalyzable portion of the pain quale may be very small, and plays a minor role in the overall experience. For example, if I burn my hand, the quality of the sensation is a composite product which includes many different sorts of response; without these, the "pure" sensation might be very meager indeed. If a different sensation, such as an itch, were imbedded in the identical context of associations, it might be interpreted as a pain. If this is so, then pain qualia as we are conscious of them can to some extent be given a functional analysis within the framework of my account of consciousness.

It seems less likely that such an analysis could be given of color qualia, but in principle it cannot be ruled out. Colors are closely associated with affective properties as are physical sensations. But the role played by color discrimination links a particular episode of color perception to past perceptions through which, in part, we identify the current object. Thus an important role played by color perception is associational: we tend to think of a given color of typical examples of certain sorts of objects. Many of these associations may be unconscious, but may nevertheless be activated during the conscious perception of color. Thus the experienced color quality may be, like pain, a composite of a "pure" sensory quality plus its context: a penumbra of associated imagery, including synesthetic associations such as the taste or feel of an orange. And as in the case of pain, the color sensation stripped of
this imagery could well be unrecognizable, and the substitution of a
different pure sensation quale within an identical context of imagery
might conceivably be unnoticed if it played the same associational role.

One might object that even so, there is a qualitative element, how­
ever thin, which cannot be accounted for in terms of its causal role.
One answer is that if the qualitative element were stripped of its associa­
tions, there might be little or no consciousness of perceiving it, since
such consciousness (introspective consciousness) entails the imagery. To
the extent that this is so, the inverted qualia problem could not arise
regarding this element in isolation, since this problem concerns the con­
sciousness of qualia. Without consciousness, a difference in two sensa­
tion qualia is simply a difference in the physiological effects of two epi­
sodes of interaction with the environment. Differences in physiological
effects are not in themselves psychological issues, and so qualia as a
problem for psychology do not arise on this level. So pure qualia as
such cannot be distinguished from their context.

However, it could also be argued that even if it is less rich than
it commonly assumed, the pure sensory quality of a sensation is experi­
enced consciously, and affects the quality of the experience in which it
occurs, so that its absence or alteration would be noticeable. It seems to
me that while there might be a noticeable difference in the case of qua­
ilia inversion, if it affected only the "core" element if the sensation qua­
ilia, it would be a difference about which little could be said or thought.
If the new quale were imbedded in the same set of associations and im­
ages, it would play the same role as the old one. Thinking about it is
nothing but activating imagery simultaneously with it (or simultaneously
with images of it), so if two qualia play the same role in arousing im­
agery, then they are thought about in the same way. Thus it seems that
any unique role played by pure, unanalyzable qualia in mental life may
be minimal.

Patricia and Paul Churchland have recently proposed a reduction­
ist account of qualia which disagrees with mine. They define qualia as
"those intrinsic or monadic properties of our sensations discriminated in
introspection", and they argue that since our sensations are (at least
token) identical with brain states, the qualia are properties of these
brain states. Thus when we are introspectively aware of qualia we are
apprehending something objective and physical in the same way as when
we apprehend properties of external objects with the outer senses. They
claim that the apparent irreducibility of qualia is the result of our
"opaque" discrimination of them, and our failure to realize that they are
the properties of brain states. Within a sophisticated theoretical frame­
work provided by a mature neuroscience, we can learn to make finer
and more accurate discriminations among and within our qualia, just as
an oenologist and a symphony conductor learn to improve their discrimi­
natory abilities within their respective areas of expertise. The Church­
lands argue that recognizing qualia as physical properties of brain
states can eliminate inverted qualia problems.

The Churchlands' position differs from mine in two important
ways. First, they hold a perceptual view of introspection, with sensa­
tions perceivable analogously to the way external objects are perceivable
with the outer senses. Second, they identify the qualia of which we are
introspectively conscious with stably recurring types of brain states,
instead of with those states plus whatever imagery and associated responses accompany such states when we are aware of them.

I have already argued against these two views about the introspective consciousness of qualia. But something more needs to be said about the Churchlands' argument. It is based upon certain analogies they claim to hold between our unsophisticated introspection of qualia and unsophisticated perception of the properties of external objects. For example, a child attending a symphony orchestra for the first time will hear the sound of the orchestra as a "seamless voice"; upon growing up and becoming a symphony conductor, she will hear the same sound as Gm7 chords and Adim chords. This is a case where the learning of a new, more sophisticated conceptual framework enables us to perceive a thing more accurately as having a complex property instead of less accurately as having a simpler property. They argue that the same thing can happen with neuroscientific sophistication, when we learn to perceive a pain as a spiking frequency in neural pathways instead of as a dull throbbing.

The Churchlands' claims seem to be aimed at the following sort of argument:

(1) If Fx is reducible to Gy, then one who perceives x as F can learn to perceive y as G.

(2) We can't learn to perceive sensation qualia as brain properties.

(3) Therefore sensation qualia are not reducible to brain properties.

The Churchlands' claim is that (2) is false, and that therefore an apparent difficulty with reductionism is not a real one.

The Churchlands clearly intend that introspecting sensations as brain properties will be experientially different from introspecting them as sensations. The change will involve more than just learning to describe our sensations in the language of neuroscience.

The genuine arrival of materialist kinematics and dynamics for psychological states and cognitive processes will constitute not a gloom in which our inner life is suppressed or eclipsed, but rather a dawning, in which its most marvelous intricacies are finally revealed—most notably, if we apply ourselves, in direct self-conscious introspection.37

The implication is that a direct awareness of, and not just an ability to describe, brain properties, will result from neuroscientific sophistication, in the same way that an awareness of the chords in the sound of the orchestra results from learning musicology. But there is an important disanalogy. Introspecting a sensation as a brain property instead of as a throbbing pain in my foot would involve an aspect shift which is not required in the orchestra example. I now experience a pain in my foot; with training, I am to learn to experience it as something in my head. The difference is like learning to see the duck instead of the rabbit in Wittgenstein's drawing. This is disanalogous to the orchestra case since there the subject does not learn to experience something other than the
original property of the sound, but rather to discriminate it in more detail. This change can be likened, not to seeing a duck instead of a rabbit, but to seeing the duck as, say, a Pekin duck in an aggressive stance. The original property, the sound as it was first heard, is composed of the chords; learning to hear them requires learning to pay more attention to the original property. But learning to introspect a pain as a brain state means learning to stop thinking about my throb­bing foot, and to think about my brain instead, and that is a major shift.

In the duck-rabbit case, the drawing was constructed expressly to make this aspect shift possible. What reason have we to think that similar shift is possible in the case of sensation qualia? The reason the Churchlands think it is possible is that they believe that sensation qualia are nothing but properties of brain states. But while that may be true, their argument about introspection was designed to support this claim, and therefore it begs the question.

It should be pointed out that in other examples the Churchlands cite as analogies to introspection of brain states, aspect shift are involved. For example, Paul Churchland talks about "the astronomer, for whom the speckled black dome of her youth has become a visible abyss", and elsewhere he mentions learning to see the earth revolving around the sun instead of the converse. However, these are examples of eliminativism rather than reductionism. It is an aspect shift to see the sky as an abyss rather than as a dome, and this fact is intimately linked with the fact that we no longer believe in a dome at all. Thus these examples will not support his claims that sensation qualia can be reduced to brain properties.

I have been suggesting that the explanation of the popularity of the perceptual model of introspection lies in the phenomenology produced by our neural mechanism of self-awareness. But this cannot be the whole story, since other accounts consistent with this phenomenology could have become entrenched. The perceptual model, and other mentalistic metaphors comprising much of folk psychology, have a cultural history: we have learned to say that there are two kinds of sense, inner and outer, as well as two kinds of act, mental and physical. There have been two interesting proposals recently concerning the origin of these theories. One is that of Eric Havelock, who has argued that the development of literacy in the oral culture of the Greeks led to beliefs in nonphysical realities and souls which could 'see' them. The oral tradition relied on the ability of heard words to arouse images which would then inspire action. Literacy allowed these words to have a visual appearance, identical in all occurrences, and attention was drawn to these written words as objects of perception apart from the images they aroused. It then appeared puzzling to Socrates and others that identical words could appear in different contexts; the puzzle was 'solved' by Plato with his postulation of the immaterial and immutable Forms, visible to the soul as the word was visible to the eye. His descriptions of the soul as an observer of nonphysical objects has passed down culturally as the root of folk psychology.

The other proposal is by Julian Jaynes, who has also suggested that consciousness was 'invented' by the Greeks as a way of accounting for the 'voices' heard by the right brain with the evolution of language in the left. They, as we, speak metaphorically of spaces and objects in
our heads, in which the "analog I" moves and observes. Like Havelock, Jaynes associates the origin of consciousness with language, but with spoken, not written language. For both, traditional theories about consciousness are elaborate constructions of a realm in which entities analogous to human bodies act on analogues of physical objects. These constructions are rooted in reality, specifically in the brain's ability to imagine or reactivate memories of physical events and bodily responses, including the perception of objects and the hearing of voices. But according to both writers, these representations are not recognized as recapitulations of bodily relations to external objects, but rather treated as an independent, nonbodily type of process, and psychological theories (folk psychology) are constructed on the basis of this assumption.

III. EMPIRICAL EVIDENCE

The theory proposed in this paper claims that the known sensory-motor system is the only one involved in mentality, and that reactivation of patterns within this system (memory traces or response patterns) is the sole source of cognition and consciousness. This claim entails neuroscientific predictions about mechanisms and types of organization in the brain, notably that there are hierarchies or "layers" of summed and abstracted representations which play a role in experience and in the initiation of action. There is much recent work on the presence of such patterns or representations in the motor cortex.

The theory makes certain psychological predictions as well as neuroscientific ones. If all thinking is activating patterns within sensory and motor systems, abstract thought processes such as problem-solving should reveal their nature in reaction-time experiments. Work by Janellen Huttenlocher supports this prediction.

A. Evidence from Neuroscience

Karl Pribram has argued that the ability of the brain to 're-present' its functions at different levels of generality and simplicity constitutes mentality. His general approach to mentality is made explicit in the case of representations governing volitional action. It is now widely believed that in such action what is represented is the use of muscles to achieve a goal, and not just the muscle movements themselves. Pribram describes and interprets work by Edward Evarts:

neurons in the motor cortex of monkeys do not fire proportionately to the amount of lengthening or shortening of a muscle involved in depressing a lever. Instead firing is proportional to the weight attached to the lever, i.e., the force necessary to move the lever. It is not the muscle or its contraction, it is the act, the use of which the muscle, is put, that is reflected in the activity of the cortical cells.

The fact that actions, not just movements of muscles, are represented in the motor cortex, has far-reaching consequences. It means that I can with my left hand write Constantinople with muscles that have never been engaged in such a performance or anything like it. . . . Thus we have at hand an explanation of the origins of the brain organization that leads to act such as moving the eyes and head.
about, the writing of plays and essays and the apparently self-generated variety of directions that the activities of men deploy. We even know a good deal about the machinery of accomplishing this sort of freedom.\textsuperscript{43}

Pribram believes that representations of action \textit{types} are stored, which makes possible novel and complex actions such as composing a play. Much research in neuroscience supports this claim.\textsuperscript{44} These representations have been independently proposed by other writers. Alvin Goldman, in "The Volitional Theory Revisited",\textsuperscript{45} cites recent support for William James’ theory that memory images of sensations produced by previous bodily movements or responses must be present for these movements to be consciously willed. Psychologist Anthony Greenwald, in work cited by Goldman, presents evidence that

an anticipatory image of feedback from an action participates in the selection and initiation of that action and that an extensive repertory of response images constitutes a "language" in which basic voluntary movements are represented.\textsuperscript{46}

This work is direct evidence only of representations of action schemata underlying voluntary actions; that these images constitute abstract thinking is at best only consistent with the evidence. The evidence is important, however, since the schematic representation of action patterns is central to my claims.

\section*{B. EVIDENCE FROM PSYCHOLOGY}

Janellen Huttenlocher performed a series of experiments on subjects asked to solve three-term series problems, such as that of identifying

the tallest of three men, John, Sam, and Tom, given that Tom is taller than Sam and John is shorter than Sam.\textsuperscript{47}

Subjects reported constructing spatial arrays, either vertical or horizontal, of the items given in the problem, and the "placing the third item with respect to these 'fixed' items according to the description in the second premise". The difficulty of constructing the spatial arrays and of placing the third item varied with the grammatical form in which the problem was presented. In particular, if in the second premise the third item (John in the above example) is the 'active' subject position, it is more easily represented as 'mobile' and hence more easily placed against the fixed array, than if it is in the 'passive' object position. Huttenlocher concludes that

The Ss easily assign spatial axes to nonspatial dimensions in these reasoning problems just as they arrange real objects along various nonspatial dimensions. The Ss report that they do not give themselves explicit instructions about the spatial positions of the items, but rather have an implicit sense of the spatial orientation of the dimension, just as when they arrange real items. . . . The Ss represent items as words (or abbreviations) rather than pictures, but they do not imagine having to write the words. Instead, each word 'ap-
pears' as E reads it, and S then treats it as a material object which can be picked up and moved.

This strategy is most economical, requiring no new procedures beyond S's ability to carry out already familiar activities in his imagination, and to use spatial images rather than real arrays to answer questions about order. By representing individual items as words, S can apply this strategy with equal ease to any ordering problems regardless of how abstract the categories being compared.48

Huttenlocher does not conclude that spatial (visual and motor) imagery is used for all abstract problem solving, but just that it is a common and useful strategy. It would require much more work to justify such a claim. Huttenlocher's work is valuable because it indicates a promising approach to this goal, and because it provides a clear example of the connection I am trying to establish between thinking and imaging.

V. SUMMARY

In this paper I have defended two main theses: (a) that we speak about mental events in metaphors whose roots are in physical event talk, and (b) mental states and events are nothing but activations of neural traces of sensori-motor events. I also claim that (b) is a major part of the explanation of (a),

The claims were defended in three stages. I began by discussing the extent of dependence of mentalistic language on bodily metaphors. I argued that although many of the physical actions referred to metaphorically are themselves goal-directed and hence seem to involve mentality, goal-directedness per se is not unique to mentality since it is characteristic of all forms of life. If our language and common-sense theories about mind are disguised metaphors for bodily actions and states, then supposedly unique mental properties such as intentionality and self-consciousness must be analyzable in terms of bodily events.

Intentionality, in Brentano's sense of "directedness upon an object" which need not really exist, can be reduced to the imagining of physical relations to physical objects. This can occur either concretely, as when one imagines an event by reactivating neural traces of similar past events, or abstractly, through the activation of abstract action patterns or schemata. This account requires that psychological events be given a 'narrow' rather than a 'wide' interpretation: intentional acts are 'in the head,' and concepts such as truth or reference are not a part of psychology.

Self-consciousness has a more complex analysis, which I approached through an examination of the theories of introspection of Armstrong and Shoemaker. Armstrong's account emphasizes the perception-like phenomenology of introspection; Shoemaker's emphasizes the difficulties for a perceptual model of introspection. Basing my account on a model of a neural mechanism for self-consciousness by Gerald Edelman, I argue that viewing self-consciousness as the imbedding of perceptions in an associative network of sensori-motor imagery captures what is of value in both Armstrong's and Shoemaker's accounts. I show how this approach can make some inroads into the problem of qualia,
and I contrast my views on qualia with those of Patricia and Paul Churchland.

I conclude with a brief discussion of some recent neuro-scientific and psychological work which lends support to some of the claims in this paper.

END NOTES

1 Part of the work on this paper was done during an NEH Summer Seminar at Cornell University in 1985. The author is grateful to its director, Sydney Shoemaker for helpful advice and suggestions.


5 Julian Jaynes has given a thorough and persuasive analysis of the metaphorical nature of mentalistic idioms, and of how mental events are analogues of physical events, in The Origin of Consciousness in the Breakdown of the Bicameral Mind (Boston: Houghton-Mifflin, 1976), Ch. 2. For related arguments about the connection between embodiment and human language see Margaret Boden, "Implications of Language Studies for Human Nature", in "Thomas W. Simon and Robert J. Scholes, eds., Language, Mind and Brain (Hillsdale, N.J.: Lawrence Erlbaum Assoc., 1982).


7 Dennett, op. cit., 8-9.


9 William James, *The Principles of Psychology*.


18 Armstrong, op. cit., 59.

19 Ibid., 64.

20 Ibid., 65.

21 Ibid., 66.


23 Ibid., 3.


25 "Introspection and the Self", 23.

26 Armstrong, Ibid., 63.


29 Ibid., 52.

30 Ibid., 76.

31 Ibid., 83.


34 Ibid., 214.

36 "Functionalism, Qualia, and Intentionality", 121.
37 "Reduction, Qualia, and the Direct Introspection of Brain States", 28.
38 Ibid., 15.
41 Jaynes, op. cit.
44 Edward Evarts, "Brain Mechanisms of Movement", op. cit.
45 In M. Brand and D. Walton, eds. Action Theory (Dordrecht, 1976), 67-84.
48 Huttenlocher, 97.