Behavior and Philosophy, *33*, 41-54 (2005).

WHAT MÜLLER'S LAW OF SPECIFIC NERVE ENERGIES SAYS ABOUT THE MIND

Howard Rachlin
Stony Brook University

ABSTRACT: Johannes Müller's law of specific nerve energies (LOSNE) states that the mind has access not to objects in the world but only to our nerves. This law implies that the contents of the mind have no qualities in common with environmental objects but serve only as arbitrary signs or markers of those objects. The present article traces the implications of LOSNE for non-physical theories of mind and for modern neural identity theory (that mental events are identical with their neurological representations) and argues that these theories are essentially inconsistent with LOSNE. Teleological behaviorism, a behavioral identity theory of the mind, identifies a person's mind with the correlation over time between that person's overt behavior and environmental objects; this behavioral conception is consistent with a revised form of LOSNE in which the mind is conceived to exist not within the body but at the borderline between the body and the world.
Key words: behaviorism, identity theory, mind, Müller, sensation, specific nerve energies

According to Boring (1957, p. 34) Johannes Müller (1801-1858) was ". . .the foremost authority on physiology of his day, and his *Handbuch*, translated immediately into English, as the primary systematic treatise." In that book Müller formulated what later came to be known as "The Law of Specific Nerve Energies" (LOSNE). It is not really a law nor was it original with Müller. Something very much like it had previously been stated by the Scottish physiologist, Sir Charles Bell; it is at least implicit in the writings of earlier physiologists and philosophers including Descartes and, to a degree, Aristotle. "The central and fundamental principle of the doctrine," according to Boring (1957, p. 82), "is that we are directly aware, not of objects, but of our nerves themselves; that is to say, the nerves are intermediaries between perceived objects and the mind and thus impose their own characteristics on the mind."

This raises the question: How do the nerves impose anything *but* their own characteristics on the mind? That is, how do the nerves tell us anything about the world? Descartes' had claimed that they don't; we already know all there is to know about the world; all the ideas we are ever going to have are already latent in our minds. Stimulation from the outside, Descartes believed, merely woke up those

AUTHOR'S NOTE: Preparation of this article was supported by a grant from The National Institutes of Health. I am grateful to William Baum and Jasper Brener for their helpful comments on an earlier draft. Please address all correspondence to Howard Rachlin, Psychology Department, Stony Brook University, Stony Brook, NY 11794-2500. Email: howard.rachlin@sunysb.edu

innate ideas. Descartes anticipated Müller's contention that the quality of the nervous message was entirely different from the quality of the stimulus:

> Nothing reaches our mind from external objects through the sense organs except certain corporeal motions [the movements of animal spirits]. . .But neither the motions themselves nor the figures arising from them are conceived by us exactly as they occur in the sense organs. . . .Hence it follows that the very ideas of the motions themselves and of the figures are innate in us. The ideas of pain, colors, sounds and the like must be all the more innate if, on the occasion of certain corporeal motions, our mind is to be capable of representing them to itself, for there is no similarity between these ideas and the corporeal motions. (*Reply to Objections II*. From Beaky & Ludlow, 1992, p. 365)

Descartes' anticipation of LOSNE was the notion that no quality of a stimulus gets past the sense organs. Nervous stimulation was like the pulling of a string within a nerve. Each stringpull opened its own valve and created a particular motion of the animal spirits ("corporeal motion"). Why is it then when one string is pulled we see a red light and when another string is pulled we see a green light? The answer, according to Descartes, is that God supplies our souls with the ideas of red and green at birth and allows each idea to be awakened only by its own particular motion. It was left for Müller to point out that even though one nerve was hooked up to the red receptors in the retina and the red receptors normally respond only to red light, anything that caused that particular nerve to fire would cause its corresponding sensation and be labeled as "red."

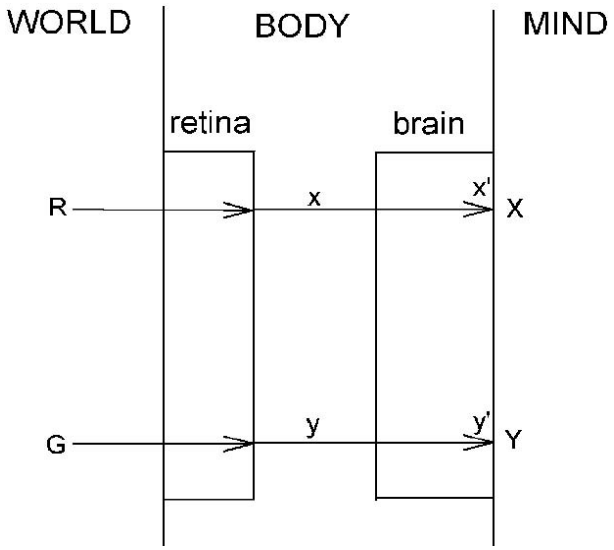

Figure 1. Müller's view of how sensations occur in the mind. Stimuli (R, G) excite the nerves (x, y) which in turn cause motions in the brain (x′, y′). These motions then cause mental states (X, Y) that have no qualities in common with the original stimuli (R, G).

Müller rejected Descartes' extreme nativism. For him, it was sufficient to say that the quality of the sensation caused by stimulation of the red-light nerve itself or, as he later speculated, the place in the brain where the red-light nerve ended, differed from the quality of the sensation caused by stimulation of the green-light nerve or the place in the brain where it ended. (Müller's later speculation, that conscious experience is determined by stimulation in a place in the brain, conforms better to modern neurophysiology; I will henceforth characterize LOSNE that way.) Figure 1 diagrams this conception. For Müller, sensations (X and Y in Figure 1) have no quality whatsoever resembling that of the physical, external red and green lights. The words "red" and "green" are merely labels we learn to attach to these different sensations. This is entirely consistent with modern thought. Our sense organs and nervous systems transform and "encode" the energies of the stimuli; our mental states are products of the transformations and encodings.

As Boring says, the crucial point of Müller's law is that our conscious experience of the stimuli is directly due to the place in the brain where the nerves end and not to the stimuli themselves. As evidence for his law, Müller cited instances where stimuli other than those acting through their appropriate sense organs nevertheless cause the nerves to fire. A blow to the head, for example might stimulate your visual nerves, in which case you would "see stars," or your auditory nerves, in which case you would "hear chimes." There are no sounds or lights within our bodies—only nervous energy. According to Müller our minds have access only to this nervous energy. From this energy, plus whatever innate tendencies our minds possess (according to Müller, the Kantian categories: space, time, moral sense, and so forth), they must construct the world. How they manage this construction became the business of all of psychology for the next hundred years and of non-behavioristic psychology even up to today.

Consider the following thought experiment. Jill is normal; when she sees a red light she feels the way we all do when we see a red light; when she sees a green light she feels the way we all do when we see a green light. But Jack is not normal; he was born with the connections between his retinas and his brain center switched; nerve-x goes from his red sensors (assuming the nervous system is that simple) to the place in his brain (y′) where nerve-y would normally go while nerve-y goes from his green sensors to the place in the brain (x′) where nerve-x would normally go. According to LOSNE, when Jill sees a green object she feels the way everyone else does when they see a green object but when Jack sees a green object he feels the way everyone else does when they see a *red* object. The mental state of seeing red for Jack would be like that of seeing green for Jill and vice-versa. Jack exists in a world where everyone else is constructed like Jill; Jack is the only one crossed up. The crucial question is this: Would Jack be hampered in any way by his unique physiology or would he get along perfectly well in the world? We may tend to think that he would at least be confused and have to learn to transform his red sensations into green behavior and vice-versa. But LOSNE implies that he would need to learn no more than Jill would.

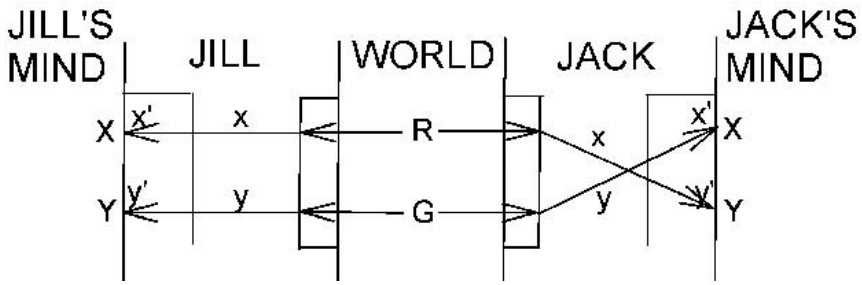

Figure 2. Given that the contents of the mind have no qualities in common with environmental stimuli (R, G), two people (Jack and Jill) with entirely opposite neural connections and opposite mental markers (X and Y) could have corresponding behaviors, hence identical mental states.

Figure 2 shows Jack along with Jill both contemplating red and green objects. Certainly, given the ultra-simplified physiology depicted in Figure 2, Jack would have no trouble communicating with Jill. Remember that, according to LOSNE, x′ has none of the properties of a red light and y' has none of the properties of a green light. The symbols x′ and y′ merely stand for different places in the brain. Therefore, unless God intervenes at some point, there can be nothing intrinsically reddish about X or greenish about Y. This was what Descartes claimed (therefore, he concluded, God must intervene). If we assume Descartes was wrong about divine intervention in human consciousness, conscious states X and Y can differ only in the sense that the letters X and Y differ—as symbols that may mean anything you want them to mean. For Jill and the rest of us X means red and Y means green while for Jack alone X means green and Y means red. It would be no more difficult for Jack to learn to call X "green" than it would be for Jill to learn to call Y "green."

Who is to say that each of us does not differ in the actual quality of our Xs and Ys? What matters, according to LOSNE, is that X and Y be different from each other and correlate, by virtue of the selectivity of our retinas, with the presence of red and green objects. Whatever associations Jill normally makes between X and other conscious states will be made by Jack between Y and other conscious states. Sensation X would be that of a warm color for Jill while sensation Y would be that of a warm color for Jack. Both sensations, whatever their differing qualities, would be associated with sensations of the sun, of fires, of blushing. Correspondingly, Y would be a cool color in Jill's mind while X would be a cool color in Jack's mind, associated, alike for Jack and Jill with sensations of trees, of plants, of the sea. At traffic lights, Jill would stop at X and go at Y while Jack would stop at Y and go at X, both behaving appropriately. The mental states X and Y would be mere markers, or labels, or signs correlated or not with other such signs.

Müller was a vitalist. He believed that the energy in the nerves has a not-purely-physical quality which enables it to travel infinitely fast from the sense organs to the mind. He believed that the quality of the sensations of red and green

light—"what it's like" to have these sensations, to use a phrase from twentieth century philosophy (Nagel, 1974)—is conveyed to the mind from the places in the brain where the nerves terminate (x′ and y′). According to Müller, the qualitative difference between X and Y serves as the basis for our conscious discrimination between red and green (the belief that it's like one thing to see red and like something else to see green) and, secondarily, as the basis for behavioral discrimination between the two stimuli.

Müller was a vitalist but his students were not vitalists. Boring says:

> In 1845. . .four young, enthusiastic and idealistic physiologists, all pupils of the great Johannes Müller, all later to be very famous, met together and formed a pact. . . .They were, in order of age, Carl Ludwig, who was then twenty-nine, Emil du Bois-Reymond, Ernst Brucke and Hermann von Helmholtz, then twenty-four. They were joining forces to fight vitalism, the view that life involves forces other than those found in the interaction of inorganic bodies. The great Johannes Müller was a vitalist, but these men were of the next generation. Du Bois and Brucke [later to become Freud's teacher] even pledged between them a solemn oath that they would establish and compel the acceptance of this truth: "No other forces than common physical chemical ones are active within the organism." (1957, p. 708)

In other words, we don't need X and Y; anything that these conscious markers can do, brain states (x′ and y′) can do as well or better.

In modern terms, Müller's students were "identity theorists." They believed that the construction of the world from nervous energy took place in the brain rather than in a nonphysical mind. Helmholtz, among his many contributions to science, went on to measure the (finite) speed of nervous transmission and to develop a theory of color vision wherein the retina analyzes colors into triplets of intensities; the three intensities are then conveyed by three nerves from the sense organ to the brain. Müller would have said that such neural triplets were synthesized into single color sensations in the mind but, for Helmholtz, whatever synthesis was required occurred wholly within the nervous system.

A great advantage of Helmholtz's identity theory as well as modern identity theory is that it recognizes the existence of unconscious mental events. For Müller, qualities X and Y were, by definition, conscious qualities; a sensation was not just a mental event it was a fundamentally conscious event. Instances where stimuli have a demonstrable effect on behavior yet are not reported as seen, as in the phenomenon of "blindsight" (Weiscrantz, 1986), would, according to Müller, be non-mental as well as non-conscious, therefore outside the purview of psychology. Identity theory neatly separates the mental from the conscious and opens up psychological investigation to methods other than conscious introspection. Psychologists and physiologists would be working in tandem—the physiologists investigating neural events on a relatively molecular level and the psychologists investigating neural events on a relatively molar level. Mental terms—sensations, perceptions, hopes, fears, dreams, thoughts, and so forth—refer, according to identity theory, to such molar neural events each in principle reducible to the

molecular events that physiologists commonly study. The project of neurocognitive psychology may be likened to the study of an unknown computer, physiologists opening it up in an attempt to discover its hardware, psychologists operating its keys and mouse and observing the results on its screen in an attempt to discover its program. On this analogy, our familiar mental terms would stand for the boxes in a flow diagram of such a program. As one modern identity theorist has pointed out (Churchland, 1986), the fact that a given computer process may be realized in many different molecular ways (analog or digital, mechanical, chemical or electronic, serial or parallel) does not make the realized program any the less physical. Temperature, for example, is differently realized in gasses, liquids and solids, yet temperature is still a physical property.

The problem with this neurocognitive identity theory, both in Helmholtz's and in the modern version, is that, despite advances in brain imaging techniques, there seems to be no set of physiological structures (molecular or molar) corresponding to each identifiable mental property. Modern sensory neurophysiology would not claim that a stimulus could be carried by a single chain of afferent nerves ending at a particular point in the brain. There are many stages on many brain levels as a stimulus affects the brain. Even a simple color is a diffuse pattern of nervous stimulation heavily influenced by contextual factors. Which one is identical with the sensation? If these factors came into focus at a "final common path" in some particular brain location you could call the focal point the sensation. But no single neural path for each sensation has been found within the nervous system. There is no place in the nervous system where the incoming stimulus stops. Sensory stimulation in its normal form runs right through the brain and out the other side so to speak without encountering anything identifiable as a sensation. The acts of a) seeing a red light and b) stopping at a red light are not two separate processes in a chain but rather a single process described in two different ways. A person who sees a red light but doesn't stop is not performing half of an act but actually seeing the red light differently from the way a person who does stop sees it. What is common among acts of discrimination of red from non-red stimuli is not a set of internal events but the set of overt actions by which the discrimination is made. In the last analysis the discrimination is the set of overt actions that correlate with the stimulus. The essential difference between a person who sees colors and one who is color blind is that, for the former, such a set exists whereas, for the latter, no such set exists.

These are old arguments (Dewey, 1896) but they undercut modern physiological identity theory.[1] If it is argued by identity theorists that sensation is identical not to stimulation in a specific place in the nervous system but to the action of a specific neural mechanism, the locus of that mechanism will have to be expanded to include the motor as well as the sensory nerves. There is no place within the body for Descartes' or Müller's or even Helmholtz's sensorium.

In other words there seems to be a fundamental contradiction within identity theory. Physiological identity theory claims that mental terms stand for events

[1] Ryle (1949) and Wittgenstein (1958, numbers 180 and 293) make similar points.

within the nervous system. Yet there is no single neural event or single locus in the brain invariably activated whenever a red stimulus, for example, affects behavior. As brain physiology has developed it has found no natural stopping point, where the input ends and the output begins, to call a sensation; there is no end point of stimulation corresponding to Descartes' pineal gland. If you start from the stimulus and work forward into the nervous system, the stimulus branches out and becomes more and more diffuse, each branch differently affected by feedback from brain structures further down the line. If you start from a given overt act and work your way backwards, sources of influence from different places in the nervous system merge like various small streams joining at different places to form a large river with back channels affecting the flow upstream. One may be tempted to say that the whole brain has the sensation. But multiple influence on behavior does not stop at the brain's border; it begins at the sense organ and ends only in the overt action itself. Rather than saying that the whole brain has the sensation it would be more correct to say that the whole person has the sensation.

One way to avoid the problem posed by the failure of modern physiology to find a sensorium is to abandon the notion that the mind can be studied by the methods of science and to avoid all talk of mental states within scientific discourse. In terms of Figure 1, nothing would seem to be lost to science by erasing everything to the right of the second vertical line. If the mind is actually *identical* to specific events in the nervous system all mental terms could be replaced by physiological terms or sets of such terms. The mind, if it exists at all, would be only an epiphenomenon. The use of mental terms in ordinary speech could be considered just the layman's way of speaking about events more precisely described in neural or neuro-cybernetic terms. This tactic would be equivalent, however, to abandoning psychology. What are cognitive psychologists studying if they are not studying memory, perception or thought? Even primarily neurological research seems to require the non-metaphorical use of mental terms (Gazzaniga, 1998).

Another conceivable solution to the problem posed by the lack of a sensorium is to seriously consider the above suggestion that sensations occur not in a specific place within an animal but in the animal's body as a whole. I call this *behavioral* identity theory. According to behavioral identity theory mental states are identical not to specific neural events but to behavioral patterns. Stopping at a red light (in the context of a pattern of discriminating red from green lights) is seen not as a two-stage process: sensing the red light and then stopping but as sensing the red light period. This is an identity theory but the identity in question is not between a person's mental processes and particular neural or neuro-cybernetic processes going on inside a person but between mental processes and the operation of the person's whole nervous system as expressed in that person's overt behavior. Instead of erasing the part of Figure 1 labeled "MIND," as Müller's students demanded, behavioral identity theory folds the figure over so "MIND" overlaps "WORLD." Behavioral identity theory preserves LOSNE in the sense that mental qualities are not isomorphic with stimulus qualities. But, according to behavioral identity theory, mental events are neither spiritual events taking place in another

world nor physiological events taking place somewhere inside the nervous system. They are rather actions of the person as a whole as expressed in overt behavior. The mental context of a given discrete act is not an internal event (physical or spiritual) efficiently causing the act. The mental context of a given act is rather the larger overt behavior pattern into which the act fits (the act's *final* cause). Behavioral identity theory, I claim, is where LOSNE, in combination with modern neurophysiology, inevitably leads.

But, it may fairly be claimed, one of the main reasons for using mental terms in the first place is to distinguish mentally controlled actions from mechanical actions. There is an essential difference between the motion of my leg in response to a hammer blow on my patellar tendon and the motion of my leg in a dance. The former is elicited mechanically and relatively simply by the hammer blow while the latter is, or at least seems to be, elicited both from inside me by my desire to dance and from outside me by the music. One of the nice things about Descartes' system is that it makes a clear distinction between these two types of action. It makes this distinction in terms of place. The place where my knee jerk originates is in the environment as the hammer hits my knee; the place where my dance step originates is in my soul, acting through my brain's pineal gland.

How may behavioral identity theory make such a distinction? One possibility is to put aside our usual way of thinking about the causes of behavior and to adopt Aristotle's teleological system. I call this system, "teleological behaviorism" (Rachlin, 1994). It essentially replaces Descartes' *spatial* distinction between mentally and mechanically controlled actions with a *temporal* distinction between these two kinds of control. As an example consider the act of stopping at a red light while driving a car. This act may be part of a temporally extended pattern of acts, generally stopping at red lights and going at green lights. That pattern may be part of a more extended pattern, generally obeying driving laws, which may be part of a still more extended pattern, generally obeying laws, which may be part of a still more extended pattern, generally obeying rules, and so forth.

Alternatively, the original act of stopping at a red light while driving a car may be seen as part of a pattern of discriminating between red and green objects, which is part of a pattern of discriminating among colors, which is part of the pattern of making sensory discriminations, and so forth. Any given act may be part of multiple sets of patterns just as a note in a symphony may be part of multiple melodies or themes.

The physiological events leading up to a given act such as stopping at a red light are its "efficient" causes. According to behavioral identity theory, mental causes are not efficient causes at all but "final" causes. The final cause of an act is the more extended pattern of acts to which it belongs; generally discriminating between red and green is the final cause of stopping at a red light in a particular case. In the context of that general discrimination, the particular case, stopping at a red light this time, is *identical* to sensing the red light. In another context (say a general failure to discriminate between red and green colors—color blindness) the very same act would not be identical to sensing the red light (the act of stopping might be a discrimination of brightness or a discrimination of place [the red light is

above the green light] or just a coincidence—the driver might have stopped to let out a passenger). The important point here is that both the final cause and the nature of the act—mental or mechanical, and, if mental, what sort—are determined by the overt act in its context of other overt acts, not by its physiological antecedents.[2]

Jack and Jill, in Figure 2, with entirely opposite neural connections (x′ and y′) would nevertheless (I assume the reader agrees) be able to coordinate their actions with respect to red and green stimuli; they both call the same stimuli red and green. According to behavioral identity theory, despite their differing physiology, Jack and Jill are both having the same sensations. Physiological identity theory would have to say the opposite—that despite their corresponding behavior Jack and Jill are having different sensations. Surely this is not a useful way to employ our mental vocabulary.

In previous articles I have discussed teleological behaviorism as it applies to self-control and social cooperation (Rachlin; 1994, 2002). In general, some particular acts are performed for their own sakes; impulsive or selfish acts such as alcoholic drinking, overeating or littering fall into this category. In modern terms, these acts are either primary reinforcers themselves or are immediately reinforced. Other particular acts would not be performed solely for their own sakes but for the sake of the wider pattern of which they are a part; self-controlled or altruistic acts such as drink-refusal, dessert-refusal, voting, or trash recycling fall into this

[2] Some modern philosophers and psychologists go part way toward this behavioral conception. O'Regan & Noë (2001), for example, argue that sensations are not internal representations but contingencies between behavior and sensory stimulation. This approach was seized upon by critics (e.g., Block, 2001) as too behavioristic; in response the authors were quick to say ". . .we are *not* behaviorists. . ." (p. 1011; italics theirs). Indeed they are not. Seeing, they concede, is not the actual sensorimotor contingency as it plays out over time between the person and the world but a wholly internal *knowledge* of that contingency. It is unclear, though, what the difference is between such knowledge and an internal representation of the contingency. (In any case this conceptual retreat is unlikely to placate philosophers like Block who believe that behaviorism is dead and want to keep it that way.) In a similar vein Bennett & Hacker (2003) present a detailed argument against the concept of internal representations. They say quite correctly that ". . .seeing the red colour of the flower does not occur in the brain at all—it occurs in the garden or the drawing room; and what is seen is a red geranium, not a sensation of red, which is neither seen nor had" (p. 133). They reject physiological identity theory: ". . .it is human beings, not their brains, that form hypotheses and make inferences" (p. 137). But Bennett and Hacker are not behaviorists either: "We are not defending a form of behaviourism" (p. 82n). Their arguments against behaviorism (e.g., p. 117) presuppose a behaviorism that, unlike teleological behaviorism, rejects mental terms. They borrow Ryle's dispositional language to talk about mental events and they are clear on what a disposition is not. It is neither physical nor mental according to them—it is not "bare bodily movement." Then what is it? Teleological behaviorism would treat dispositions as patterns of overt behavior over time rather than internal states (like defining probabilities as relative frequencies over time rather than internal states). Bennett and Hacker, eager to distance themselves from behaviorism, never consider this conception.

category. Such acts may not be immediately reinforced but are part of a highly valuable pattern of acts (sobriety, healthy life style, group cohesiveness). Aristotle (in *De Anima* and the *Nicomachean Ethics*) showed that the language of mental terms (sensation, perception, imagination, thought, as well as social responsibility and happiness) applies naturally to such behavioral patterns (see Rachlin, 1994, for an extended discussion of Aristotle's psychology and ethics).

This is not to say that modern physiological and cognitive investigations are any less valuable or important than behavioral investigation—on the contrary. The point of this article is just that LOSNE implies that mental states may be studied directly by investigation of behavioral patterns; physiological investigations and many cognitive investigations are directed to the mechanisms underlying those states (their efficient causes). Moreover, many cognitively oriented psychology experiments, especially in the areas of judgment, decision and choice, may be interpreted in terms of behavioral patterning as well as in terms of underlying mechanisms (Rachlin, 1989).

In contradiction to behavioral identity theory, common sense might claim, A: *We just know* that our sensations are inside us; therefore Jack's and Jill's sensations must differ even though their actions correspond. But common sense would also claim (in agreement with Descartes), B: Our sensations are non-arbitrarily related to objects in the world; they are not mere mental markers or arbitrary neural encodings.

The usual cognitive, neurological, or spiritual interpretation of LOSNE preserves claim-A with respect to sensations but sacrifices claim-B. This is a bad bargain. It drains all utility from the concept of sensation and makes it into an epiphenomenon, intrinsically private and unrelated to the world. Given claim-A and LOSNE, a scientific psychophysics would be impossible. Verbal reports of mental events (introspection) would be completely unreliable since the very objects of introspection might differ from person to person yet be described (as by Jack and Jill) in identical terms or might be identical yet be described differently. For this reason, Skinner (1938) as well as some modern philosophers (Quine, 1960, for example) denied the utility of mental terms in psychology.

But it is possible to retain the utility of mental terms in psychology by dropping claim-A and resuscitating something like claim-B. The behavioral interpretation of sensations (as particular discriminatory acts in the context of a general discrimination—a set of overt actions correlated over time with their objects) retains a non-arbitrary, functional, relation of sensations with the world (as well as their poetic use); the relation between an object in the world and the sensation of that object is, according to behavioral identity theory, the relation of a stimulus and its function as a guide to behavior.

Let us now consider an objection that may be raised to the thought experiment illustrated in Figure 2. The objection might run as follows:

> You deliberately chose red and green lights to illustrate LOSNE because these stimuli lack emotional valence. Colors serve in the world (as red and green traffic lights do) not so much for their own qualities as for the

> qualities they signal. Since the stimuli you chose are themselves signals, of course their neural representations are also signals and could stand for anything. Jack and Jill communicate with each other exactly to the extent that red and green colors are unsought (or unavoided) for their own sake. If the spiritual or neural representations of the colors were pleasant or unpleasant in themselves, Jack would quickly discover that he was abnormal.

Note how much this objection already concedes to behavioral identity theory (that sensations consist of overt classificatory behavior of whole organisms). All environmental events that serve primarily as signals for other events, including virtually all the language we hear and read, can be internally represented only as arbitrary (spiritual or physiological) signals. This includes not only abstract colors and sounds but also whole objects and complexes of objects (perceptions). Consider for example our fathers and mothers. Everything that was said about red and green lights as they affect Jack's and Jill's nervous systems could be said about their mother and father (Jack and Jill are brother and sister, it may now be revealed). If the image of their mother were represented in Jack's brain in the place corresponding to the image of their father in Jill's brain, and vice-versa, Jack and Jill would still be able to communicate perfectly with each other about their parents—regardless of how important they were to them. According to LOSNE, Jack and Jill may have different feelings for their common mother but these different feelings could not consist only of differing spiritual or neural representations. Conversely, Jack and Jill could have the same feelings for their mother but differing spiritual or neural representations of her. Why? Because it is their mother whom they love or hate, not the representations—which (according to LOSNE) have none of the qualities of their mother. What does it mean then for two people to have different feelings about a common object? For behavioral identity theory, it can mean only that they behave differently with respect to that object. "What it's like" to love or hate your mother is to behave in one way or another toward her over an extended period of time.

There is nothing in this behavioral conception of the mind that says discrimination must occur at a single instant. Jack may love his mother and Jill may hate her at this moment yet they may both be behaving, at this moment, in the same way (they may both be kissing her—one on each cheek). Discrimination (even the simple kind between red and green lights) is the correlation of behavior with environmental events over time. Strictly speaking, discrimination cannot happen at a moment. However, it is a common verbal convention to refer to a temporally extended event as if it did exist at a moment. Two violinists, for example, may each be playing a different melody *at this moment* even though, at this moment, they are playing the very same note. But this is a verbal convention referring to what the violinists are actually doing and has nothing to do with what may be going on within their heads.

According to behavioral identity theory, to love your mother or anyone else is not to have your heart go pitter-patter at the sight of her but to behave toward her

in a loving way over a long period of time. False love is a discrepancy not of your behavior with your heart or of your behavior with your mental representation but of your immediate behavior with your behavior in the long run. If Jack loves his mother while Jill hates her and they are both kissing her right now the kiss is like a common note in two different melodies being played by two different violinists. The kiss is but a single act common to two different behavioral patterns. The love and the hate are in the patterns.

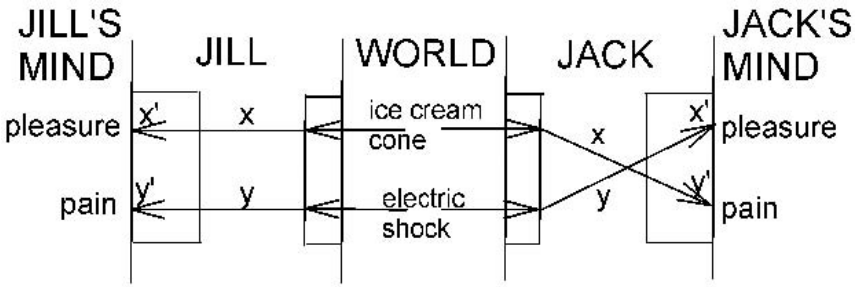

Figure 3. Like Figure 2 except the stimuli are pleasures and pains rather than red and green objects. Here the behaviors in response to the stimuli (not shown) are innately different hence Jack's and Jill's mental states are different.

What then about pleasure and pain themselves? Figure 3 repeats Figure 2 but instead of red and green lights substitutes a typically pleasurable stimulus (an ice cream cone) and a typically painful stimulus (an electric shock). Suppose that Jack is crossed up by some bad fairy with respect to these stimuli so that whatever causes pleasure for the rest of us causes pain for Jack while whatever causes pain for the rest of us causes pleasure for Jack. Under such conditions Jack would be truly mixed up. He would be punished for eating ice cream cones and rewarded for sticking forks into electric sockets. To survive he would have to learn to avoid whatever gave him pleasure and to approach whatever gave him pain. It would conceivably be possible to teach him to overcome these tendencies and to communicate with normal people about objects in the world but he would at the least be a very unfortunate person.

Of what does his misfortune consist? To approach this question let us consider still another transformation of Jack's nervous system. As we said, the bad fairy has arranged, before his birth, that little Jack will hate licking ice cream cones and love sticking forks into electric sockets. But, in the nick of time, after this transformation, and before Jack was actually born, a good fairy has come along and altered his reflexes so that he now has an innate tendency (he can't help it) to approach whatever is painful to him and to avoid whatever is pleasurable. He has the same (spiritual or physiological) pains as before but can't help eagerly approaching whatever causes them (his mother's breasts at first, ice cream cones later); the greater the pain intensity the more eager is his approach. Correspondingly, he has (spiritual or physiological) pleasures but can't help avoiding whatever causes them (shocks, fires, blows on the head). The more

intense the pleasure, the louder are his screams, the more fierce his struggles to escape. In other words, Jack behaves normally—just like Jill—except his spiritual or neural representations of pleasurable and painful objects are opposite to hers. If he were to deny his pleasure in the objects he seeks no one would believe him. "I hate this ice cream cone," he would say as he hungrily devours it. But he cannot even verbally deny his pleasure or signal his displeasure with the ice cream by grimacing; doing so would be to signal to others that he would rather not partake of this common pleasure; others may tend to comply by withholding it. To grimace would be to avoid, at least indirectly, those inner pains that he is compelled to seek.

Some readers, I think, will resist this second transformation. Unlike the first, it will seem almost inconceivable in a way that the first one (with red and green lights), though no less impossible, was not. The reason why it is inconceivable (or just barely conceivable) for Jack to approach what is truly painful and avoid what is truly pleasurable is that, for all our intents and purposes, true pleasure and true pain are identical, not to spiritual or neural events, but to the behavior of approach and avoidance—not to momentary approach and avoidance but to consistent approach and avoidance over time. Jack may seek normally painful stimuli for some greater good, but if a long-term observer (Jill, for instance) sees no greater good and if the behavior is frequent, Jill will judge that Jack really likes to do it.

What makes pleasure and pain different from sensations of red and green is not the greater subjectivity of the former pair of mental events but the greater specificity of behavior in response to pleasurable and painful stimuli than in response to red and green stimuli. A newborn infant responds to his mother's breast by seeking the nipple and sucking; when he is slapped, he cries. But the infant has to learn what to do in response to red and green signals.

As previously said, a behavioral identity theory of the mind would fold Figures 1, 2, and 3 so that the areas labeled MIND overlap those labeled WORLD. For the behaviorist, our minds, including our pleasures, pains, sensations of red and green, and perceptions of our parents, exist not in another, spiritual, sphere different from the physical world, not in one or another piece of our brains, but in our habitual interactions with the world.[3] Otherwise the conception of Jack's double transformation holds; we would have to imagine the possibility of a person like Jack, behaving just like everyone else from birth to death but fundamentally, radically, different in his mind. Of all the implications of LOSNE only the behavioral one avoids this violation of common sense.

[3] This is not to deny that there are pleasure centers and pain centers in the brain (although I do not believe that the evidence for them is conclusive). Rather, I claim, pain itself is not identical with stimulation of what is normally a pain center and pleasure itself is not identical with stimulation of a pleasure center. Pain and pleasure centers are perhaps essential parts of the mechanisms underlying pain and pleasure (like the engine is an essential part of the mechanism underlying a car's acceleration) but not pains and pleasures themselves (not the acceleration itself). Pain itself is the consistent avoidance and pleasure itself the consistent seeking of their objects (see Rachlin, 1985, for an extended discussion of pain in behavioral terms).

References

Beakley, B., & Ludlow, P. (1992). *The philosophy of mind: Classical problems/contemporary issues*. Cambridge, MA: MIT Press.

Bennett, M. R., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford: Blackwell.

Block, N. (2001). Behaviorism revisited. *Behavioral And Brain Sciences*, *24*, 997.

Boring, E. G. (1957). *A history of experimental psychology* (2nd ed.) New York: Appleton-Century-Crofts.

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

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review*, *3*, 357-370.

Gazzaniga, M. S. (1998). *The mind's past*. Berkeley: University of California Press.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*, 435-450.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral And Brain Sciences*, *24*, 939-1031.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Rachlin, H. (1985). Pain and behavior. *The Behavioral And Brain Sciences*, *8*, 43-52.

Rachlin, H. (1989). *Judgment, decision, and choice*. New York: W. H. Freeman.

Rachlin, H. (1994). *Behavior and mind: The roots of modern psychology*. New York: Oxford University Press.

Rachlin, H. (2002). Altruism and selfishness. *The Behavioral And Brain Sciences*, *25*, 239-296.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson House.

Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton Century.

Weiscrantz, L. (1986). *Blindsight*. Oxford: Oxford University Press.

Wittgenstein, L. (1958). *Philosophical investigations* (3rd ed.) G. E. M. Anscombe (Trans.) New York: Macmillan.