

Behaviorism, mentalism, and “our ordinary conception of ourselves”

*Conductismo, mentalismo y “nuestra concepción
ordinaria de nosotros mismos”*

**Robert L. Arrington
Georgia State University**

Abstract

Behaviorists often view their theories as being in conflict with and overturning essential aspects of our ordinary understanding of ourselves. They take this ordinary conception to be dualistic in nature, incorporating numerous kinds of mental events and forms of mental causation. I argue that these behaviorists mischaracterize the ordinary conception. Following Wittgenstein, I attempt to provide a more adequate characterization of it with respect to such matters as voluntary behavior, intentional action, the expression of feelings, emotions, and thoughts, and the identification of the self. I also urge that Wittgenstein's notion of an expression yields a richer concept of behavior than the one often employed by behaviorists. The latter, I claim, is simply the dualist's concept of behavior with the mental dimension deleted. Behaviorists are, as it were, one-armed dualists.

Key Words: Wittgenstein, behaviorism, dualism, mentalism, intentional action, voluntary behavior, the will, the self, expression, feelings, thoughts.

Resumen

Los conductistas a menudo visualizan sus teorías en conflicto con y trastornando aspectos esenciales de nuestra comprensión ordinaria de nosotros mismos. Asumen que esta concepción ordinaria es de naturaleza dualista, incorporando numerosas clases de eventos mentales y formas de causación mental. Sostengo

que estos conductistas mal representan a la concepción ordinaria. Siguiendo a Wittgenstein, intento proporcionar una caracterización más adecuada de ello respecto a asuntos tales como la conducta voluntaria, la acción intencional, la expresión de sentimientos, emociones y pensamientos, y la identificación del yo mismo. También reclamo que la noción de Wittgenstein de una expresión constituye un concepto más enriquecedor de la conducta que aquel empleado a menudo por los conductistas. Este último, sostengo, es simplemente el concepto dualista de la conducta con la dimensión mental borrada. Los conductistas son, como si fueran, dualistas de un solo brazo.

Palabras clave: Wittgenstein, conductismo, mentalismo, acción intencional, conducta voluntaria, la voluntad, el yo mismo, expresión, sentimientos, pensamientos.

Behaviorists have not been reluctant to point out what they see as the philosophical implications of their doctrine. Schwartz and Lacey, in their book *Behaviorism, Science, and Human Nature*,¹ refer to behavior theory as a "world view," and they indicate two beliefs they consider central to it. First, "Behavior theorists claim that the causes of behavior lie not within the actor but in the environment" (p.14), and second, "The central thesis of behavior theory is that virtually all significant voluntary human actions can be understood in terms of their past relations to rewards and punishments" (p.15). They proceed to contrast this behaviorist conception of human nature with the conception most people have of it:

It is hard to imagine a view of human nature more opposed to our ordinary conception of ourselves than this one. We have all grown accustomed to thinking of ourselves as the controllers of our own lives. Society holds us responsible for our actions, and we readily accept that responsibility. We place an extraordinary value on our freedom of choice, and we resent and resist any efforts at coercion. We formulate our own goals, and feel we act on the basis of our own preferences and desires, and not in keeping with external pressures. All of these features of our conception of ourselves are built into our everyday understanding of human action in terms of goals and purposes. (pp.15-16)

Schwartz and Lacey point to ways in which behavior theory rejects this everyday conception:

1. Barry Schwartz and Hugh Lacey, *Behaviorism, Science, and Human Nature*, New York and London: W.W. Norton and Company, 1982.

Behavior theory challenges this conception of ourselves. If you want to know why someone did something, do not ask. Analyze the person's immediate environment until you find the reward. If you want to change someone's actions, do not reason or persuade. Find the reward and eliminate it. The idea that people are autonomous and possess within themselves the power and the reasons for making decisions has no place in behavior theory. (p.16)

This behaviorist view, Schwartz and Lacey admit, will be found "repugnant and degrading" by many people, but they hint that behavior theory, and presumably its world view, is becoming "an increasingly influential part of our culture" (ibid).

As we proceed, let us keep in mind the main tenets of the "ordinary conception of ourselves" rejected by behaviorism, at least according to Schwartz and Lacey:

1. We are the controllers of our own lives;
2. We are held responsible for our actions;
3. We value our freedom of choice, which, presumably, we have;
4. We often act on the basis of our self-formulated goals and our preferences and desires, as opposed to external pressures;
5. We have within us the power to make decisions, which we make on the basis of reasons;
6. We can tell someone else why we did something;
7. We can reason with and persuade others to act differently;
8. We are autonomous agents.

These characteristics are cardinal features of what elsewhere I have called, following Wilfred Sellars, the manifest image of human nature.² Talking of the manifest image is another way of speaking of "our ordinary conception of ourselves". For economy of expression, we can therefore say that Schwartz and Lacey object to the manifest image and would like to replace it with a scientific, behaviorist image.

In his book *Behaviorism: A Conceptual Reconstruction*, G.E. Zuriff³ defines behaviorism in ways quite similar to those put forth by Schwartz and Lacey. Zuriff approvingly quotes Skinner on the notion of a proper theory of behavior: "it must abolish the conception of the individual as a doer, as an originator of action" (Zuriff, p.176). This leads Zuriff to say "The behaviorist conception thus

2. Robert L. Arrington, "The Rationality Machine," Second International Congress on Behaviorism and the Science of Behavior, Palermo, Sicily, 1994, unpublished. See Wilfred Sellars, *Science, Perception and Reality*, New York: The Humanities Press, 1963.

3. G.E. Zuriff, *Behaviorism: A Conceptual Reconstruction*, New York: Columbia University Press, 1985.

differs from the popular notion of persons as active agents who are responsible for their own acts. This agent, or self, although basic to our everyday understanding of the world, finds no place in the behaviorist conceptual framework" (ibid.). One problem that Zuriff notes with the concept of agency is that it implies free will (ibid.), a notion incompatible with "the deterministic conception of the universe underlying the physical sciences" (ibid.). This incompatibility presumably results from the fact that an agent with free will initiates actions that cannot be understood in terms of the laws of science (the laws of physics, or, if you please, the laws of behaviorism). This suggests to Zuriff that the private volitions of the free agent bring about, cause, actions. "If the choices of the agent are not determinately linked to observable external causes, then the only secure bond is between the private actions of the agent and subsequent behavior" (Zuriff, p.177). Zuriff disapproves of this conception of agency on the grounds that "the causes of behavior cannot be known by public observation but only by the introspections of the agent" (ibid.). Introspection, he claims, is "unscientific and nonobjective" (ibid.). Moreover, he rejects the conception of the self that is associated with this notion of (free) agency: "the self is generally understood as the seat of conscious awareness which directs behavior and is distinct from the material processes by which behavior is manifested" (ibid.). This immediately suggests, he warns, "a dualism of mind and matter" (ibid.).

This, then, is the picture we get from Zuriff. Behaviorism stands in opposition to the conceptual scheme basic to our everyday understanding of the world, i.e. the manifest image, in which an agent, through inner volitions, initiates actions. Behaviorists reject the notion that we explain actions by tracing them to the causal activity of the agent —such explanations "merely assign to a mysterious entity, the self, whatever properties are required to account for an otherwise inexplicable action" (Zuriff, p.178). Zuriff even thinks that explanations in terms of agent causation amount to a form of animism in which "human actions are understood as the outward manifestation of the internal workings of a spirit, soul, or homunculus" (ibid.). Behaviorism, on the contrary, stands firm in the view that all behavior can be understood in terms of external causes and the laws of behavior linking those external causes to types of behavior. Thus it rejects dualism and animism, and with them the notion of a mysterious, action-causing agent.

If one thinks of the human being in terms of the kind of Cartesian dualism Zuriff has in mind, many of the elements of the everyday conception of human nature identified by Schwartz and Lacey as well as by Zuriff himself seem to fall on the mind or spirit side of the mind/body dichotomy, viz. choices, decisions,

goals, preferences, desires, reasons for acting, responsibility, free will, etc. Likewise for these writers, many of the characteristics of a good behavioral theory appear to fall on the body side of dualism: external pressures, environmental causes, physical determinism, scientific lawfulness, public observability. In fact, what intrigues me about the way our apologists for behaviorism set up the debate between behaviorism and our "ordinary conception of ourselves" is the set of dualistic assumptions embodied in their descriptions both of the "ordinary conception" and behaviorism. Behaviorists, I would like to say, are dualists in spite of themselves, or, at least, one-armed dualists. What I mean by this is that dualism defines the terms in which behaviorism characterizes itself and the terms of behaviorism's debate with the manifest image. Deliberations, decisions, choices and the like are described as mentalistic events or processes that are publicly unobservable. The behaviorist accepts the dualistic mentalist's conception of mental events —*that's what they would be like, if they existed* or if a behavioral science were interested in them. The only things left over when one subtracts the mental half of reality are the motions of physical objects and their causal, mechanistic interactions. Among these motions and interactions we must find the subject matter of behaviorism. Behavior will have to be seen as bodily motions, and the laws governing these motions will be laws connecting the motions to their external causes. Although some behaviorists have explicitly embraced this conception of their subject matter, their practice and their preaching, as Vickie Lee has reminded us,⁴ are seldom consistent. Certainly when behavioristic studies are done of human beings, actions, not motions, are often the subject matter. This is even true of experiments involving animals, as when the behavior of pressing a button is investigated. Pressing a button is something the rat, pigeon, or human being *does*. Pressing a button as an action is distinct from the motion of the foot or fingers coming into contact with the button —a distinction Wittgenstein was getting at when he asked what the difference is between my raising my arm and my arm rising (*PI*, §621).⁵ Actions may involve bodily motions, but they cannot be reduced to such.

Many behaviorists are reluctant, however, to talk of more than bodily motions because such discourse seems to involve mentalistic intentions, acts of will, thoughts and deliberations. Surely, as scientists, behaviorists cannot allow these ghostly events into their descriptions and theories! But what if intentions,

4. Vickie L. Lee, *Beyond Behaviorism*, Hillsdale, N.J.: Erlbaum Associates, 1988, pp. 38ff.; 95-96.

5. Ludwig Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe, London: Macmillan, 1953. References to *Philosophical Investigations* (*PI*) are to numbered sections unless otherwise indicated.

thoughts and deliberations —as we ordinarily talk of them— are inaccurately described in the terms of dualistic mentalism? What if these psychological characteristics have behavioral criteria, so that we cannot even understand what we are talking about in referring to them if we drop the references to behavior? The philosopher Wittgenstein has argued, as I am sure you know, that dualistic mentalism is a conceptually confused way of talking about psychological matters. In Wittgenstein's hands, what I have been calling the manifest image of human beings comes to be seen as thoroughly behavioral, not dualistic or mentalistic *in a dualist manner*. Wittgenstein sums up this image in his striking remark "The human body is the best picture of the human soul" (*PI*, p. 178).

Let me give you an example of the way in which Wittgenstein seeks to overcome the dualism and opposition between the notion of the mental and that of the physical. He often focuses on *expressions* (*Äußerungen*) of feelings. Take, for instance, a grimace, or a person trembling with rage. The grimace is not just the outward manifestation of an inner feeling —you can, as it were, see the pain in the grimace ("there was a lot of pain in his face," we might say). Likewise, when a person trembles with rage, the rage is not hidden from view, unobservable by all except the person who is in a rage. The trembling *and* the rage are palpable —the rage is palpable in the trembling. As Wittgenstein puts it with regard to expressions of pain, "If I see someone writhing in pain with evident cause I do not think: all the same, his feelings are hidden from me" (*PI*, p. 223). The feelings are there, *in the writhing*. And as regards the psychological state of being "out of humor", Wittgenstein has this to say: "'I noticed that he was out of humour.' Is this a report about his behavior or his state of mind? . . . Both; not side-by-side, however, but about the one *via* the other" (*PI*, p.179). When we describe behavioral expressions like trembling with rage, writhing in pain, being out of humor and the like, we are describing mental states expressed *in* behavior. The range of such expressions is vast. As Wittgenstein puts it, "Look into someone else's face, and see consciousness in it, and a particular *shade* of consciousness. You see *on it, in it* (my emphasis), joy, indifference, interest, excitement, torpor and so on" (*Z*, §220).⁶

Of course, a person may tremble with fear as well as with anger. Do we observe the trembling, and infer the fear or the rage? When we observe the trembling, do we ask, is this the outward manifestation of fear or rage? No, we directly observe the person trembling with rage, which is quite different from

6. Ludwig Wittgenstein, *Zettel*, trans. G.E.M. Anscombe, Berkeley: University of California Press, 1970. All references to *Zettel* (*Z*) are to numbered sections.

seeing someone trembling with fear. The "physiognomy" of the two forms of behavior, as Wittgenstein would put it, is different. Wittgenstein expresses the matter like this:

We do not see facial contortions and make inferences from them (like a doctor framing a diagnosis) to joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give any other description of the features. Grief, one would like to say, is personified in the face.

This belongs to the concept of emotion. (*Z* §225)

The last remark is immensely important. Our very concept of grief includes its standard behavioral manifestations. We *mean by* grief something having these expressions. Grief is not, then, something essentially inner and hidden from public view.

The attempt to divide up phenomena like trembling with rage into two components, one hidden and mental, the other observable and bodily, derives from the confused image of the human person we get from dualistic mentalism. This metaphysics misses altogether our ordinary notion of an expression of emotion or feeling, a behavior in which feeling or emotion is made manifest, which is to say, is given shape and intensity, and a thousand other nuances, through the behavior of the person. In defining itself in opposition to the mentalism of dualism, behaviorism risks missing altogether the immensely rich domain of behavioral expressions. This is unfortunate. It maligns the everyday conception of ourselves (the manifest image) by attributing to it unsavory spiritualistic elements that are in fact foreign to it; and it restricts behaviorism to a set of conceptions which yields a highly impoverished understanding of behavior.

A very important subset of *Äußerungen* are verbal expressions. Just as we express our feelings through such things as grimaces and scowls, we can express them by telling others how we feel. Wittgenstein has suggested that the proper way to understand the verbal expression of pain—"I'm in pain"—is to see it as a form of pain-behavior that replaces the natural cry of pain (*PI*, §244; also see *RPP* I, §151⁷ and *Z* §545). It is instructive to think of other first-person, present-tense psychological statements as also being instances of expressive behavior, even though there may be no instinctive behavior that they replace. For instance, we can verbally express our thoughts: we do so all the time with such

7. Ludwig Wittgenstein, *Remarks on the Philosophy of Psychology*, Vol. 1, Oxford: Blackwell, 1980 and Chicago: University of Chicago Press, 1980. All references to *Remarks on the Philosophy of Psychology*, Vol. 1 (*RPP* I) are to numbered sections.

verbal idioms as “I think the Democrats will win the next election” and “I strongly believe the Democrats have the higher moral ground.” Cartesian dualists at work in the field of philosophical psychology will be inclined to take these verbal expressions as *reports*, informative descriptions of the inner thoughts and beliefs of the subject. After all, the verbal expression is a physical, bodily thing, the thoughts and beliefs something different altogether, something mental. Behaviorists will be inclined to interpret these reports as the result of some inner observations of one’s thoughts and beliefs, and insofar as the subject matter of these observations (these introspections) is not available for public view, the behaviorist is suspicious of such reports. But both the mentalist and the behaviorist fail to see that an expression of thought is an embodiment of thought—I tell you what I think; a politician puts forward his beliefs for public debate. There is nothing to the thoughts or beliefs that does not make its way into the sincere verbal expression of them. If I *think* the Democrats will win next year and *say* “I think the Democrats will win next year” I have totally, fully expressed this thought. As Wittgenstein notes, I don’t know my thought better than my words represent it (*RPP* I, §576). I’m sure to have other thoughts associated with any given thought, but I could express them as well. Both the mentalist and the behaviorist fail to acknowledge Wittgenstein’s admonition: “Ask not: ‘What goes on in us when we are certain that . . .?’—but: How is ‘the certainty that this is the case’ manifested in human action?” (*PI*, p.225). Ask not: “What goes on in us when we think?”—but: How is ‘the thought that this is the case’ manifested in human action?

But, the objection will be forthcoming from dualists of both mentalistic and behaviorist conviction, a person can think without saying anything, just as a person can feel pain without grimacing or showing any other sign of it. Likewise, a grimace might be insincere, there being no pain behind it; the expression of a thought might be duplicitous, the actual thought being the reverse of what is said. So how can the thinking, or the pain, be embodied in the verbal expression of it?

It is true, of course, that there are malingerers and insincere politicians. But note that we are occasionally taken in because they are very good at dissimulation—their behavior convinces us of their thoughts and feelings. Also note that we sometimes find them out, and we do so by spotting inconsistencies among their various actions. Never do we confirm a case of a person being in pain by feeling that person’s pain, and we never spot the malingerer by somehow detecting the internal absence of inner pain in the person. We denounce a particular expression of thought as insincere by contrasting it with other things

the person says and does. In other words, while there are sincere and insincere expressions of our "states of mind", we draw the distinction in terms of patterns of behavior. As Wittgenstein remarks, pretending is a "very special pattern in the weave of our lives" (*PI*, p.229). And a sincere expression of pain, or thought, is another pattern and weave. The sincere behavior embodies pain or thought; the insincere behavior embodies pretense, which, we must remember, is also a psychological state.

Wittgenstein has taught us to be careful with first-person present tense psychological utterances like "I have a headache" or "I believe the Democrats will win the next election." Dualists take them to be reports or descriptions of inner observations, but Wittgenstein questions this notion of an inner observation. "If you observe your own grief, which senses do you use to observe it? A particular sense; one that *feels* grief?" (*PI*, p.187). In what sense is observing grief different from simply feeling it? Moreover, the notion of *observation* makes no sense here. Observing something involves, at least potentially, putting oneself in a position to detect it (*PI*, p.187) —but one doesn't "put oneself in a position" to observe grief, or pain, or joy, or any other psychological state. As to someone who suggests that first-person, present tense psychological utterances are descriptions and not expressions, Wittgenstein remarks: "let him ask himself how anyone learns to describe a table and how he learns to describe his own thoughts. And that only means: let him look and see how one judges the description of a table as right or wrong, and how the description of thoughts..." (*RPP I*, §572). We look again at the table, to check our description. Do we look again at our thoughts? Perhaps from a different angle? Finally, if first-person, present tense psychological utterances *were* (introspective) reports or descriptions, they could be correct or incorrect, and the person making the report could be making a mistake instead of getting things right. But it makes no sense for me to say that I may be wrong about how I feel or what I think. In saying "I have a headache" or "I think the Democrats will win", I may be insincere but not mistaken. Indeed, if it makes no sense to speak of a person being in error about his feelings and thoughts, neither does it make any sense to speak of his *knowing* what these feelings and thoughts are. Thus there is no introspection constituting my source of information about what I feel or think. The concern about introspection, and the development of an epistemology designed to avoid it, is based upon a mythical, mentalistic conception of first-person, present-tense psychological statements. Instead of talking about people having mental occurrences inside them and knowing about them or making mistakes about them through introspection, let us say something like this: people have feelings and thoughts

and they express them in various forms of behavior; knowing what another person is feeling is a matter of observing and interpreting his expressive behavior; the person himself *has* the feelings or thoughts—he doesn't need to *know* them. *Nothing* gets between an emotion (or thought) and its expression.

In most of his discussions of psychological statements and expressions, Wittgenstein is questioning the "inner/outer" distinction that is the hallmark of dualism. The dualist, be he an unreconstructed mentalist or a reconstructed behaviorist, takes behavior to be something outer. The mentalist thinks of it as the outer manifestation of an inner mental event; the behaviorist rejects the inner dimension as mysterious, mythical, and unscientific but still interprets behavior as a mentalist would. But in the case of *Äußerungen*, we have forms of behavior that "reek" with mind. There is no inner/outer distinction to be drawn here, precisely because in these expressions we have the breakdown of the distinction. Mind and body are both real, but they are not distinct and in opposition to one another. I think Wittgenstein sums up his view nicely in the following passage: "What is fear? What does 'being afraid' mean? If I wanted to define it at a *single* showing—I should *play-act* fear" (*PI*, p.188). The fact that one could define fear by play-acting it, by behaving in a certain way, shows that fear is no essentially inner experience, just as it shows that behavior is no essentially, and mere, outer event.

If Wittgenstein is correct in his descriptions of the manifest image of human nature—as I think he is—then in talking about this image we should avoid characterizing it in dualistic, mentalistic terms—as Schwartz and Lacey, Zuriff, and many other behaviorists do. And we should avoid defining behaviorism in opposition to the "ordinary conception of ourselves" when this ordinary conception is construed dualistically. To do so would be to define it in relation to a straw man, which makes it a less than interesting theory.

To provide further defense of these claims, let us return to some of the things attributed by behaviorists to the manifest image—things these behaviorists reject. First of all, there is the notion of a self that supposedly causes actions. Behaviorists repudiate such a self, claiming that all actions are caused by external, environmental events. To fill in the picture a bit more, when acting voluntarily the mentalistic self as agent presumably engages in acts of will, and these acts of will in turn cause his actions. Is this an accurate characterization of the manifest image? To get at the conceptual scheme involved in our everyday conception of things, we need to investigate the ways in which we talk about these matters. Do we ever say anything like "he caused an act of willing" or "his act of willing caused him to go to the grocery store"? Such locutions are, in fact,

monstrous. We do, of course, say that a person did something voluntarily, or involuntarily, but it is only a dualist who will interpret such talk in terms of agents causing acts of will which in turn cause actions. If we investigate what Wittgenstein calls the "grammar" of our talk of voluntary behavior, it will turn out that one can understand it perfectly well in terms that have nothing to do with mysterious self-causation and acts of will.

In fact, the notion of a self causing acts of will that in turn cause other acts is incoherent, as Gilbert Ryle showed us long ago.⁸ When the self causes an act of will, does it do so voluntarily or involuntarily? Surely not involuntarily! But if voluntarily, would this not require another act of will as a cause—an act of will to bring about an act of will!? Presumably an act is voluntary by being caused by an act of will. And is this prior act of will (the one that occurs when we will to will) itself voluntary? The notion that a voluntary act is one preceded by an act of will leads to an infinite regress of acts of will—which means: it leads to nonsense. As Wittgenstein puts it, properly understood "'Willing' is not the name of an action; and so not the name of any voluntary action either" (*PI*, §613). We have misunderstood the ordinary concept of the will if we interpret it as something that produces actions. Again Wittgenstein: "One produces a sneeze or a fit of coughing in oneself, but not a voluntary movement. And the will does not produce sneezing, nor yet walking" (*Z* §579).

To be sure, our everyday discourse is full of statements and claims like "Joe did x" or "Sam is doing y". That is to say, we readily attribute *agency* to people, and the notion of agency does incorporate the idea that we sometimes behave voluntarily. I shall later return to this notion of voluntary behavior, but for the moment let us look at Zuriff's suggestion that agency involves the activity of a mysterious "self". In saying that Joe stole the beer, am I making reference to Joe as a mysterious "self" hidden from the view of ordinary observation? Of course not. Our criteria of identity for a person or agent are a bit vague and loose, but this does not prevent us in most cases from determining that it was Joe and not, say, Sam who stole the beer. Criteria of identity for persons focus primarily on physical appearance and bodily continuity, supplemented on occasion with psychological characteristics. Physical appearance is important but sometimes not sufficient. If Sam is Joe's identical twin, we might have to put Joe's body at the scene of the crime—or put Sam's body elsewhere—to ascertain that it was Joe and not Sam who did it. If for some reason we can't do this, we must appeal

8. Gilbert Ryle, *The Concept of Mind*, London: Hutchinson and Co., and New York: Barnes and Noble, 1949, p. 67.

to the character traits and the like that distinguish Joe from Sam, and these traits can be given in terms of their behavioral criteria. In making such identifications, we are not peddling any heavy metaphysics of the self, nor do we need to appeal to data that a behaviorist would find suspicious.

But what about a person's intentions? Surely they make a difference to whether a person did something voluntarily, or to whether we hold him responsible for what he did. Aren't these intentions private, inaccessible, and faintly malodorous? Insofar as the manifest image involves the attribution of intentions in explaining behavior, must we not reject it? The manifest image surely sanctions talk about intentions, but before we conclude that we must reject it for this reason, let us recall what I said earlier about behavioral *expressions*. Just as we express our feelings and thoughts, we express our intentions—we tell others what we intend to do. Of course, sometimes we lie and our expressions mislead others. But sometimes we are caught in our lies, just as sometimes—often, in fact—we are truthful in expressing our intentions. Our judgments on what a person intended to do are judgments based on observations of behavior, including verbal behavior and the various contexts of behavior. Nothing mysterious is involved, and there is no metaphysical or epistemological obstacle to our knowing the intentions of others. Which is not to say that ascertaining someone's intentions is an easy matter—nothing could be further from the truth. What is needed for the task, however, is the shrewd, insightful observer of human behavior, someone with a broad knowledge of mankind (*PI*, p. 227) and what Wittgenstein calls a "nose" for the "imponderable evidence" that includes "subtleties of glance, of gesture, of tone" (*PI*, p.228). Perhaps a psychologist (perhaps a behaviorist) would be good at this!

Are we the controllers of our lives? What could it mean to ask this question other than to inquire if we often do what we want and intend to do. And surely we almost always do what we intend to do. Do we ever act on our own preferences, desires, goals, and reasons? It would be perverse to think that we do not, for that would mean that every time we acted we would have done something we didn't want to do, something contrary to our preferences and goals, and this is simply not true. Some of us, sometimes, are forced to do things we don't like to do, and some unfortunate citizens of countries in the grip of tyranny or despotism often find themselves in this predicament. But in places where there is considerable political and social liberty, controlling one's own destiny is the common lot of people.

But, we might be told, even when one's actions are consistent with one's intentions and desires, these intentions and desires do not cause them. The

causation is rather to be located in the external environment. If we want to explain why a person acted as he did, we must see his action as a function of the contingencies involved in his previous behavior. This behaviorist rebuttal begs a number of questions. First, it assumes that the manifest image incorporates the idea that when a person acts in accordance with his desires or intentions, these latter "mental" states *cause* his behavior. The behaviorist then denies this mental causation and locates it elsewhere. But do we ever say that a person's intention caused him to act? We say that he acted intentionally—or unintentionally—but that is arguably a different matter from being caused to act by one's intention. Do we ever say that someone's feelings or desires caused him to do something? Yes we do, but let us ask ourselves precisely when, in ordinary life, we would say such a thing. We might say this in the case of a person whose desires overwhelmed him—when he was a victim, say, of obsessive love or jealousy—or when his emotions were so strong he couldn't control them or when he gave into them. As we all know, these things do happen—but they do not always happen, and in fact they are fairly infrequent.

In the normal case, our desires give us a *reason* to do what we do. A person normally acts because he has certain desires and certain beliefs about how to attain the objects of these desires. These beliefs and desires constitute his reasons for action, not its causes. As Norman Malcolm has noted, we do not speak of a person's act as the *effect* of his desires and beliefs.⁹ Rather, the agent claims that his act is *justified* by the desires and beliefs he has. We may agree or disagree with this judgment, but even when we disagree, our own judgment is that the agent acted from *bad* reasons, not that his act resulted from his beliefs and desires as causes. So the behaviorist who takes issue with the manifest image concerning the locus of causality for actions has set up a straw man as his opponent. With the exception of abnormal cases in which desires or emotions do cause us to act, our desires are not—in the manifest image—conceived to be causes of actions at all.

The behaviorist also begs the question in claiming that we can explain and understand an action only if we find its causes in the external environment. He approaches this issue with a peculiar conception of "understanding" and "explanation" in mind—needless to say, a "causal" conception. In his eyes, to explain why a person did something is to give a functional account of his behavior in which the dependent variable is the behavior and the independent variables are the factors involved in the person's history of reinforcement. But there is

9. Norman Malcolm, *Consciousness and Causality* (with D.M. Armstrong), Oxford: Blackwell, 1984, p. 72.

another kind of explanation we have access to—at least we do if we operate within the manifest image. We explain an act by reference to the agent's reasons for actions. Although some philosophers, following Donald Davidson,¹⁰ think that reasons are a form of cause, it can be shown, I think, that this is not the case. Reasons have a totally different kind of relationship to an act than causes do. Reasons must entail, via a practical syllogism, that an act has a certain desirability characteristic. Causes don't entail anything.

It is important to note that explanations via reasons are often available to us, and when they are, they give us a potentially complete understanding of why the agent acted as he did. We understand why someone went to the grocery store if we know that he wanted to buy milk. We understand why a person goes to college when we are informed that he thought doing so was the best way to become what he wanted to become—say a successful businessman. Sometimes, to be sure, we won't quite understand an action even when we know the beliefs and desires that constituted the reasons for it. This will happen when these beliefs and desires are unusual or odd, and then we will want to know why the person believed *this* or wanted *that*. Further reasons are then sought which explain why he had the belief or the desire. These further reasons are also justificatory in nature; they justify the initial belief/desire, showing it to be rational in light of the other beliefs and desires of the agent. When they are available to us, reason explanations, either at the basic level or at higher levels, give us an understanding, and a complete understanding, of an action. Sometimes they are not available to us; sometimes we can't understand how a rational agent could have believed such and such, or wanted such and such. In cases like these, we may conclude that the agent did not in fact act rationally, and it is precisely in these circumstances that we may say his behavior was caused—by his obsessive desires or his strange, delusory ideas, or something else. But in most cases, we don't need causal explanations and wouldn't be helped by them—we have perfectly adequate, and complete, rational explanations for a person's behavior.

A behaviorist might benefit from noting the resemblance between the concept of rational behavior and the concept of operant behavior. Vickie Lee has described operant behavior as means-end behavior, a matter of engaging in a certain activity in order to obtain a certain consequence.¹¹ That, of course, is

10. Donald Davidson, "Actions, Reasons, and Causes," in his *Essays on Actions and Events*, Oxford and New York: Oxford University Press, 1982, pp. 3-19.

11. Vickie Lee, *op cit.*, p.77.

precisely what we often do when we engage in rational behavior. Our reasons for acting, in such a case, relate to the end or consequence we shall obtain in doing so. The *prima facie* resemblance between operant and rational behavior should lead us to ask if a behaviorism committed to the notion of an operant is in fact so far removed from the manifest image of human behavior.

We must, of course, be careful here. Operant behaviorists talk of operant *conditioning*, and this makes it clear that they are operating from within a causal perspective. Moreover, much of the operant behavior they study lies below the threshold of (apparent) rationality: the organism is either incapable of giving reasons for its behavior or in fact does not do so. This implies that not all means-end behavior is rational in nature, and we must not wish to suggest that it is.

Moreover, the behaviorist who is able to explain a piece of non-rational means-end behavior may wish to infer from this fact that he has the explanatory devices necessary to explain *all* means-end behavior, including what we, using the manifest image, call rational behavior. This inference may lead to a form of reductionism, the claim that rational behavior is just a specific kind of means-end behavior, all of which can be explained in terms of the functional laws of operant psychology. This would require that the deliberative, reason-based behavior we have talked so much about is itself governed by the functional principles of operant conditioning. And it would imply that we are under an illusion when we speak of acting from reasons and acting voluntarily —the facts are, the behaviorist would claim, that we are always controlled, not by our beliefs and desires, however behavioristically construed, but by the external environment.

Before the behaviorist reduction of rational behavior to causal sequences can become at all plausible, however, the reduction must be able to account for the variety of distinctions that surround talk of rational behavior. For instance, rational behavior contrasts with irrational behavior; behavior, although rational, may be based on factual errors; desires may conflict with one another, leading to rankings of desires; weakness of the will may lead one to do what in fact one knows one ought not to do. Above all, any reduction of rational behavior to causal sequences must account for the normativity involved in rational behavior. How statistical regularities add up to the oughts, duties, rights, and goods of rational behavior is not easy to see.

Finally, let us take a brief look at the free will issue. Behaviorists reject free will, thinking it unscientific, incompatible with the lawfulness of human behavior, and inconsistent with universal determinism. Is the manifest image commit-

ted to this type of freedom? I have already questioned the notion that agents cause acts of will that in turn cause actions, and, it seems to me, it is this notion that behaviorists usually have in mind when they reject free will. If I am correct, no such agent causation and volitional causation is at work in the manifest image. But how does our "ordinary conception of ourselves" view behavior that is free? First of all, the locution that has priority in the manifest image will be 'voluntary behavior', and so we need to ask after the concept of voluntary behavior. One very tempting answer is that provided by philosophers we call compatibilists. They accept the principle of determinism and hence see all behavior as caused. But they draw a distinction between behaviors having external causes (being pushed or blown by the wind) and those having internal causes (certain desires and beliefs on the part of the agent). If a person's normal beliefs and desires cause the act, then, according to the compatibilist, the person acts voluntarily or freely. If, that is to say, his action is the product of his normal character or personality, he acts voluntarily. When coerced by external forces to do something he does not wish to do, or by internal forces to do something he would not do given his normal character or desires, he acts involuntarily. Many distinctions must be drawn in order to make the compatibilist's case plausible. It must be shown how we are to distinguish a person's normal character or desires from abnormal ones, how we are to handle the case of threats, and how we are to avoid the counterclaim that even normal desires are caused by things in the external environment. My problem with compatibilism, however, is that it is inconsistent with things I have said about the manifest image. According to this conception of things, I have suggested, desires and beliefs usually don't *cause* actions; rather, they justify them. Thus I cannot accept the compatibilist's claim that free action is simply an action having a particular *kind* of cause. How, then, are we to understand voluntary behavior?

As usual, Wittgenstein is helpful here. In his late writings on the philosophy of psychology he has this to say: "There is a peculiar play of movements, words, facial expressions, etc., as of expressions of reluctance, or of readiness, which characterize the voluntary movements of the normal human being" (*RPP I*, §841). Moreover, voluntary behavior is behavior of which it makes sense to say that one does it gladly, or that one decides to do it, or that it was done because of certain considerations (*ibid.*). A movement of my body is involuntary if I, the agent, don't know that it is taking place (*RPP I*, §844), or if it makes no sense for someone to command me to do that sort of thing. These remarks set us off in the right direction, although they hardly give us anything approaching an adequate account. I myself would like to highlight the notion of rational behav-

ior as an important ingredient in our concept of voluntary behavior. Rational actions, actions that are justified in the agent's mind by reasons, are paradigm cases of voluntary actions. These are actions arising from decisions that reflect prior deliberation and practical argumentation. These paradigms do not exhaust the field, for there are many voluntary actions that do not involve prior deliberation or decision. These latter instances must be considered case by case, noting their similarities to the paradigms and their differences as well. In doing this, we delineate a broad, hopefully not overly intellectualized class of voluntary actions. Obviously I cannot undertake here the task of constructing this class. All I will say is this. The class of voluntary actions is not to be understood in terms of the dualistic, mentalistic conception of it. It involves none of the absurd acts of will or unmotivated acts of a mentalistic self that dualists *and* behaviorists refer to. In debating the question of human freedom, the behaviorist needs to join in a debate with those who equally repudiate the mentalistic conception of freedom. He needs to enter a debate with the true representatives of "our ordinary conception of ourselves."¹²

12. I wish to thank Stuart Shanker for the conversation that inspired this essay. Spirited discussions with Jack Marr helped shape the direction of its argument.