PHILOSOPHY OF MIND

THIRD EDITION

JAEGWON KIM

A Member of the Perseus Books Group

Westview Press was founded in 1975 in Boulder, Colorado, by notable publisher and intellectual Fred Praeger. Westview Press continues to publish scholarly titles and high-quality undergraduate- and graduate-level textbooks in core social science disciplines. With books developed, written, and edited with the needs of serious nonfiction readers, professors, and students in mind, Westview Press honors its long history of publishing books that matter.

Published by Westview Press,
A Member of the Perseus Books Group

Find us on the World Wide Web at www.westviewpress.com.

Westview Press books are available at special discounts for bulk purchases in the United States by corporations, institutions, and other organizations. For more information, please contact the Special Markets Department at the Perseus Books Group, 2300 Chestnut Street, Suite 200, Philadelphia, PA 19103, or call (800) 810-4145, ext. 5000, or e-mail special.markets@perseusbooks.com.

Designed by Trish Wilkinson
Set in 10.5 point Minion Pro

Library of Congress Cataloging-in-Publication Data

Kim, Jaegwon.
Philosophy of mind / Jaegwon Kim.—3rd ed.
p. cm.
ISBN 978-0-8133-4458-4 (alk. paper)
1. Philosophy of mind. I. Title.
BD418.3.K54 2011
128'.2—dc22
E-book ISBN 978-0-8133-4520-8 2010040944

10 9 8 7 6 5 4 3 2 1

CHAPTER 3

Mind and Behavior

Behaviorism

Behaviorism arose early in the twentieth century as a doctrine on the nature and methodology of psychology, in reaction to what some psychologists took to be the subjective and unscientific character of introspectionist psychology. In his classic *Principles of Psychology*, published in 1890, William James, who had a major role in establishing psychology as a scientific field, begins with an unambiguous statement of the scope of psychology:

> Psychology is the Science of Mental Life, both of its phenomena and of their conditions. The phenomena are such things as we call feelings, desires, cognitions, reasonings, decisions, and the like.[1]

For James, then, psychology was the scientific study of mental phenomena, with the study of conscious mental processes as its core task. As for the method of investigation of these processes, James writes: "Introspective observation is what we have to rely on first and foremost and always."[2]

Compare this with the declaration in 1913 by J. B. Watson, who is considered the founder of the behaviorist movement: "Psychology . . . is a purely objective experimental branch of natural science. Its theoretical goal is the prediction and control of behavior."[3]

This view of psychology as an experimental study of publicly observable human and animal behavior, not of inner mental life observed through private introspection, dominated scientific psychology and associated fields until the 1960s and made "behavioral science" a preferred name for psychology in universities and research centers around the world, especially in North America.

The rise of behaviorism and the influential position it attained was no fluke. Even James saw the importance of behavior to mentality; in *The Principles of Psychology*, he also writes:

> *The pursuance of future ends and the choice of means for their attainment are thus the mark and criterion of the presence of mentality* in a phenomenon. We all use this test to discriminate between an intelligent and a mechanical performance. We impute no mentality to sticks and stones, because they never seem to move for *the sake of* anything.[4]

It is agreed on all sides that behavior is intimately related to mentality. Obviously, what we do is inseparably connected with what we think and want, how we feel, and what we intend to accomplish. Our behavior is a natural expression of our beliefs and desires, feelings and emotions, and goals and aspirations. But what precisely is the relationship? Does behavior merely serve, as James seems to be suggesting, as an *indication*, or a *sign*, that a mind is present? And if behavior is a sign of mentality, what makes it so? If something serves as a sign of something else, there must be an underlying relationship that explains why the first can serve as a sign of the second. Fall in the barometric pressure is a sign of an oncoming rain; that is based on observed regular sequences. Is behavior related to minds in a similar way? Not likely: You can wait and see if rain comes; you presumably can't look inside another mind to see if it's really there!

Or is the relationship between behavior and mentality a more intimate one? Philosophical behaviorism takes behavior as *constitutive* of mentality: Having a mind just *is* a matter of exhibiting, or having a *propensity* or *capacity* to exhibit, appropriate patterns of behavior. Although behaviorism, in both its scientific and philosophical forms, has lost the sweeping influence it once enjoyed, it is a doctrine that we need to understand in some depth and detail, since not only does it form the historical backdrop of much of the subsequent thinking about the mind, but its influence lingers on and can be discerned in some important current philosophical positions. In addition, a proper appreciation of its motivation and arguments will help us gain a better understanding of the relationship between behavior and mentality. As we will see, it cannot be denied that behavior has something crucial to do with minds, although this relationship may not have been correctly conceived by behaviorism. Further, reflections on the issues that motivated behaviorism can help us gain an informed perspective on the nature and status of psychology and cognitive science.

The Cartesian Theater and the "Beetle in the Box"

On the traditional conception of mind deriving from Descartes, the mind is a private inner stage, aptly called the Cartesian theater by some philosophers,[5] on which mental actions take place. It is the arena in which our thoughts, bodily sensations, perceptual sensings, volitions, emotions, and all the rest make their appearances, play out their assigned roles, and then fade away. All this for an audience of one: One and only one person has a view of the stage, and no one else is permitted a look. Moreover, that single person, who "owns" the theater, has a full and authoritative view of what goes in the theater: Nothing that appears on the stage escapes her notice. She is in total cognitive charge of her theater. In contrast, the outsiders must depend on what she says and does to guess what might be happening in the theater; no direct viewing is allowed.

I know, directly and authoritatively, that I am having a pain in my bleeding finger. You can see the bleeding finger, and hear my words "Oh damn! This hurts!" and come to believe that I must be experiencing a bad pain. Your knowledge of my pain is based on observation and evidence, though probably not explicit inference, whereas my knowledge of it is direct and immediate. You see your roommate leaving the apartment with her raincoat on and carrying an umbrella, and you reason that she thinks it is going to rain. But she knows what she thinks without having to observe what she is doing with her raincoat; she knows it directly. Or so it seems. Evidently, all this points to an asymmetry between the first person and the third person where knowledge of mental states is concerned: Our knowledge of our own current mental states is *direct*, in that it is not mediated by evidence or inference, and *authoritative*, or *privileged*, in that in normal circumstances, it is immune to the third person's challenge, "How do you know?" This question is a demand for evidence for your knowledge claim. Since your knowledge is not based on evidence, or inference from evidence, there is nothing for you to say, except perhaps "I just know."

Early in the twentieth century, however, some philosophers and psychologists began to question this traditional conception of mentality; they thought that it led to unacceptable consequences, consequences that seemingly contradict our ordinary assumptions and practices involving knowledge of other minds and our use of language to talk about mental states, both ours and others'.

The difficulty is *not* that such knowledge, based as it is only on "outer" signs, is liable to error and cannot attain the kind of certainty with which we supposedly know our own minds. The problem, as some saw it, goes deeper:

It makes knowledge of other minds not possible at all! Take a standard case of inductive inference—inference based on premises that are less than logically conclusive—such as this: You find your roommate listening to the weather report on the radio, which is predicting heavy showers later in the day, and say to yourself, "She is going to be looking for her umbrella!" This inference is liable to error: Perhaps she misunderstood the weather report or wasn't paying attention, or she rather enjoys getting wet. Now compare this with our inference of a person's pain from her "pain behavior." There is this difference: In the former case, you can check by further observation whether your inference was correct (you can wait and see whether she looks for her umbrella), but with the latter, further observation yields only more observation of her behavior, never an observation of her pain! Only she can experience her pains; all you can do is to see what she does and says. And what she *says* is only behavior of another kind. (Maybe she is very stoic and reserved about little pains and aches.) One hallmark of induction is that inductive predictions can be confirmed or disconfirmed—you just wait and see whether the predicted outcome occurs. For this reason, inductive procedures are said to be self-correcting; predictive successes, or lack thereof, are their essential constraint. In contrast, predictions of inner mental events on behavioral evidence cannot be verified one way or the other, and not subject to correction. As a result, there is no predictive constraint on them. This makes it dubious whether these are legitimate inferences from behavior to inner mental states at all.

The point is driven home by Ludwig Wittgenstein's parable of "the beetle in the box." Wittgenstein writes:

> Suppose everyone had a box with something in it; we call it a "beetle." No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at *his* beetle. Here it would be quite possible for everyone to have something different in his box.[6]

As it happens, you have a beetle in your box, and everyone else says that they too have a beetle in their box. But what can you know from their utterances, "I have a beetle in my box"? How would you know what they mean by the word "beetle"?

The apparent answer is that there is no way for you to know what others mean by "beetle," or to confirm whether they have in their boxes what you have in yours: For all you know, some may have a butterfly, some may have a little rock, and perhaps others have nothing at all in their boxes. Nor can others know what you mean when they hear you say, "I have a beetle in my box."

As Wittgenstein says, the thing in the box "cancels out whatever it is." It is difficult to see how the word "beetle" can have a common meaning that can be shared by speakers, or how the word "beetle" could have a role in the exchange of information.

A deeper lesson of Wittgenstein's beetle, therefore, is that it is mysterious how, on the Cartesian conception of the mind, we could ever fix the meaning of the word "pain" and use utterances like "I have a pain in my knee" to impart information to other speakers. For the pain case seems exactly analogous to the beetle in the box: Suppose you and your friends take a fall while running on the track and all of you bruise your knees. Everyone cries out "My knee hurts!" On the Cartesian picture, something is going on in each person's mind, but each can observe only what's going on in her mind, not what's going on in anyone else's. Is there any reason to think that there is something common, some identical sensory experience, going on in everyone's mind, in each Cartesian theater? Pain in the mind seems just as elusive as the beetle in the box. You are experiencing pain; another person could be feeling an itch in the knee; still others could have a tickle; some may be having a sensation unlike anything you have ever experienced; and some may not be having any sensation at all. As Wittgenstein would have said, the thing in each mind cancels out whatever it is.

Evidently, however, we use utterances like "My knee hurts" to communicate information to other people, and expressions like "pain" and "the thought that it's going to rain" have intersubjective meanings, meanings that can be shared by different speakers. Your pain gets worse and you decide to go to a clinic. Gently tapping your kneecap with her fingers, your physician asks, "Does it hurt?" You reply, "Yes, it does, Doctor." This is a familiar kind of exchange in a medical office, and it can be important to diagnosis and treatment. But the exchange makes no sense unless the words "the knee hurts" on your doctor's mouth mean the same as "the knee hurts" on your mouth; unless the expression has a shared meaning for you and your doctor, your reply could not count as an answer to your doctor's question. You and your doctor would be talking past each other. Our psychological language, the language in which we talk about sensations, likes and dislikes, hopes and regrets, thoughts, emotions, and the rest, is an essential vehicle of social interchange and interaction; without a language in which we communicate with each other about such matters, social life as we know it is scarcely imaginable. For this to be possible, the expressions of this language must have by and large stable and invariant meanings from speaker to speaker. What we have seen is that the privacy of the Cartesian minds may well infect psychological language, making it essentially private as

well. The problem is that a private language fails as a genuine language, because the defining function of language is to serve as an instrument of interpersonal communication. All this seems to discredit the Cartesian picture of the mind as an inner theater for an audience of one.

Behaviorism is a response to these seemingly unacceptable consequences of the Cartesian conception of the mind. It rejects the traditional picture of how our mental expressions acquire their meanings by referring to private inner episodes, and attempts to ground their meanings in publicly accessible and verifiable facts and conditions about people. According to the behaviorist approach, the meanings of mental expressions, such as "pain" and "thought," are to be explained by reference to facts about observable behavior—how people who have pain or thoughts act and behave. But what is meant by "behavior"?

What Is Behavior?

As our first pass, we can take "behavior" to mean whatever people or organisms, or even mechanical systems, *do* that is *publicly observable*. "Doing" is to be distinguished from "having something done," though this distinction is not always clear. If you grasp my arm and pull it up, the rising of my arm is not something I do; it is not my behavior (but your pulling up my arm is behavior—your behavior). It is not something that a psychologist would be interested in investigating. But if I raise my arm—that is, if I cause it to rise—then it is something I do, and it counts as my behavior. It is not assumed here that the doing must in some sense be "intentional" or done for a purpose; it is only required that it is proximately caused by some occurrence internal to the behaving system. If a robot moves toward a table and picks up a book, its movements are part of its behavior, regardless of whether the robot "knows" or "intends" what it is doing. If a bullet punctures the robot's skin, that is not part of its behavior, not something it does; it is only something that happens to it.[7]

What are some examples of things that humans and other behaving organisms do? Let us consider the following four possible types:

i. *Physiological reactions and responses*: for example, perspiration, salivation, coughing, increase in the pulse rate, rising blood pressure.[8]
ii. *Bodily movements*: for example, walking, running, raising a hand, opening a door, throwing a baseball, a cat scratching at the door, a rat turning left in a T-maze.

iii. *Actions involving bodily motions*: for example, greeting a friend, writing an e-mail, going shopping, writing a check, attending a concert.
iv. *Actions not involving overt bodily motions*: for example, judging, reasoning, guessing, calculating, deciding, intending.

Behaviors falling under (iv), sometimes called "mental acts," evidently involve "inner" events that cannot be said to be publicly observable, and behaviorists do not consider them "behavior" in their sense. (This, however, does not necessarily rule out behavioral interpretations of these activities.) Those falling under (iii), although they involve bodily movements, also have clear and substantial psychological components. Consider the act of writing a check: Only if you have certain cognitive capacities, beliefs, desires, and an understanding of relevant social institutions can you write a check. You must have a desire to make a payment and the belief that writing a check is a means toward that end. You must also have some understanding of exchange of money for goods and services and the institution of banking. The main point is this: A person whose observable behavior is indistinguishable from yours when you are writing a check is not necessarily writing a check, and a person who is waving his hand just like you are waving yours may not be greeting a friend although you are (try to think how these things can happen). Something like this is true of other examples listed under (iii), and this means that none of these count as behavior for the behaviorist. Remember: Public observability is key to the behaviorist conception of behavior. This implies that if two behaviors are observationally indistinguishable, they must count as the "same" behavior.

So only those behaviors under (i) and (ii) on our list—what some behaviorists called "motions and noises"—meet the behaviorist requirements. In much behaviorist literature, there is an assumption that only physiological responses and bodily motions that are in a broad sense "overt" and "external" are to count as behavior. This could rule out events and processes occurring in the internal organs; thus, internal physiological states, including states of the brain, would not, on this view, count as behavior, although they are physical states and conditions that are intersubjectively accessible. The main point to remember, though, is that however the domain of behavior is circumscribed, behavior is taken to be bodily events and conditions that are publicly accessible to all competent observers. Behavior in this sense does not enjoy the kind of privileged access granted to the first person in the Cartesian picture. That is, *equal access for all* is of the essence of behavior as conceived by the behaviorist.

Logical Behaviorism: A Positivist Argument

Writing in 1935, Carl G. Hempel, a leading logical positivist, said, "We see clearly that the meaning of a psychological statement consists solely in the function of abbreviating the description of certain modes of physical response characteristic of the bodies of men and animals."[9]

This is what is called "logical behaviorism," because it is based on the supposed close logical connections between psychological expressions and expressions referring to behavior. It is also called "analytical behaviorism" or "philosophical behaviorism" (to be distinguished from scientific, or methodological behaviorism; see below). Fundamentally, it is a claim about the translatability of psychological sentences into sentences that ostensibly refer to no inner psychological occurrences but only to publicly observable aspects of the subject's behavior and physical conditions. More formally, the claim can be stated like this:

> *Logical Behaviorism I*. Any meaningful psychological statement, that is, a statement purportedly describing a mental phenomenon, can be *translated*, without loss of content, into a cluster of statements solely about behavioral and physical phenomena.

And the claim can be formulated somewhat more broadly as a thesis about the behavioral definability of all meaningful psychological expressions:

> *Logical Behaviorism II*. Every meaningful psychological expression can be *defined* solely in terms of behavioral and physical expressions, that is, expressions referring to behavioral and physical phenomena.

Here "definition" is to be understood in the following fairly strict sense: If an expression E is defined as E^*, then E and E^* must be either synonymous or conceptually equivalent (that is, as a matter of meaning, there is no conceivable situation to which one of the expressions applies but the other does not).[10] Assuming translation to involve synonymy or at least conceptual equivalence, we can see that logical behaviorism (II) entails logical behaviorism (I).

Why should anyone accept logical behaviorism? The following argument extracted from Hempel represents one important line of thinking that led to the behaviorist position:

1. The meaning of a sentence is given by the conditions that must be verified to obtain if the sentence is true (we may call these "verification conditions").
2. If a sentence has a meaning that can be shared by different speakers, its verification conditions must be accessible to each speaker—that is, they must be publicly observable.
3. Only behavioral and physical phenomena (including physiological occurrences) are publicly observable.
4. Therefore, the sharable meaning of any psychological sentence must be specifiable by statements of publicly observable verification conditions, that is, statements describing behavioral and physical conditions that must hold if the psychological statement is true.

Premise (1) is called "the verifiability criterion of meaning," a central doctrine of the philosophical movement of the early twentieth century known as logical positivism. The idea that meanings are verification conditions is no longer widely accepted, though it is by no means dead. However, we can see and appreciate the motivation to go for something like the intersubjective verifiability requirement in the following way. We want our psychological statements to have public, sharable meanings and to serve as vehicles of interpersonal communication. Suppose someone asserts a sentence S. For me to understand what S means, I must know what state of affairs is represented by S (for example, whether S represents snow's being white or the sky's being blue). But for me to know what state of affairs this is, it must be one that is accessible to me; it must be the kind of thing that I could in principle determine to obtain or not to obtain. It follows that if the meaning of S—namely, the state of affairs that S represents—is to be intersubjectively sharable, it must be specified by conditions that are intersubjectively accessible. Therefore, if psychological statements and expressions are to be part of public language suitable for intersubjective communication, their meanings must be governed by publicly accessible criteria, and only behavioral and physical conditions qualify as such criteria. And if anyone insists that there are inner subjective criteria for psychological expressions as well, we should reply, the behaviorist would argue, that even if such existed, they (like Wittgenstein's beetles) could not be part of the meanings that can be understood and shared by different persons. Summarizing all this, we could say: Insofar as psychological expressions have interpersonal meanings, they must be definable in terms of behavioral and physical expressions.

A Behavioral Translation of "Paul Has a Toothache"

As an example of behavioral and physical translation of psychological statements, let us see how Hempel proposes to translate "Paul has a toothache" in behavioral terms. His translation consists of the following five clauses:[11]

a. Paul weeps and makes gestures of such and such kinds.
b. At the question "What is the matter?" Paul utters the words, "I have a toothache."
c. Closer examination reveals a decayed tooth with exposed pulp.
d. Paul's blood pressure, digestive processes, the speed of his reactions, show such and such changes.
e. Such and such processes occur in Paul's central nervous system.

Hempel suggests that we regard this list as open-ended; there may be many other such "test sentences" that would help to verify the statement that Paul is having a toothache. But how plausible is the claim that these sentences together constitute a behavioral-physical translation of "Paul has a toothache"?

It is clear that as long as translation is required to preserve "meaning" in the ordinary sense, we must disqualify (d) and (e): It is not a condition on the mastery of the meaning of "toothache" that we know anything about blood pressure, reaction times, and conditions of the nervous system. Even (c) is questionable: Why can't someone experience toothache (that is, have a "toothachy" pain) without having a decayed tooth or in fact any tooth at all? (Think about "phantom pains" in an amputated limb.) (If "toothache" means "pain caused by an abnormal physical condition of a tooth," then "toothache" is no longer a purely psychological expression.) This leaves us with (a) and (b).

Consider (b): It associates *verbal behavior* with toothache. Unquestionably, verbal reports play an important role in our finding out what other people are thinking and feeling, and we might think that verbal reports, and verbal behavior in general, are observable behavior that we can depend on for knowledge of other minds. But there is a problem: Verbal behavior is not pure physical behavior, behavior narrowly so called. In fact, it can be seen that verbal behavior, such as responding to a question with an utterance like "I have a toothache," presupposes much that is robustly psychological; it is a behavior of kind (iv) distinguished earlier. For Paul's response to be relevant here, he must *understand* the question "What is the matter?" and *intend to express the belief* that he has a toothache, by uttering the sentence "I have a toothache."

Understanding a language and using it for interpersonal communication is a sophisticated, highly complex cognitive ability, not something we can subsume under "motions and noises." Moreover, given that Paul is having a toothache, he responds in the way indicated in (b) *only if he wants to tell the truth*. But "want" is a psychological term, and building this clause into (b) would again compromise its behavioral-physical character. We must conclude that (b) is not an eligible behavioral-physical "test sentence." We return to some of these issues in the next section.

Difficulties with Behavioral Definitions

Let us consider beliefs: How might we define "S believes that there are no native leopards in North America" in terms of S's behavior? Pains are associated with a rough but distinctive range of behavior patterns, such as winces, groans, screams, characteristic ways in which we favor the affected bodily parts, and so on, which we may collectively call "pain behavior" (recall Hempel's condition [a]). However, it is much more difficult to associate higher cognitive states with specific patterns of behavior. Is there even a loosely definable range of bodily behavior that is characteristically and typically exhibited by all people who believe that there are no native leopards in North America, or that free press is essential to democracy? Surely the idea of looking for bodily behaviors correlated with these beliefs makes little sense.

This is why it is tempting, perhaps necessary, to resort to the idea of *verbal behavior*—the disposition to produce appropriate verbal responses when prompted in certain ways. A person who believes that there are no native leopards in North America has a certain linguistic disposition—for example, he would tend to utter the sentence "There are no native leopards in North America," or its synonymous variants, under certain conditions. This leads to the following schematic definition:

> S believes that $p =_{\text{def}}$ If S is asked, "Is it true that *p*?" S will answer, "Yes, it is true that *p*."

The right-hand side of this formula (the "definiens") states a *dispositional* property (*disposition* for short) of S: S has a disposition, or propensity, to produce behavior of an appropriate sort under specified conditions. It is in this sense that properties like being soluble in water or being magnetic are called dispositions: Water-soluble things dissolve when immersed in water, and magnetic objects attract iron filings that are placed nearby. To be soluble at time *t*, it

need not be dissolving at *t*, or ever. To have the belief that *p* at time *t*, you only need to be disposed, at *t*, to respond appropriately if prompted in certain ways; you need not actually produce any of the specified responses at *t*.

There is no question that something like the above definition plays a role in finding out what other people believe. And it should be possible to formulate similar definitions for other propositional attitudes, like desiring and hoping. The importance of verbal behavior in the ascription of beliefs can be seen when we reflect on the fact that we are willing to ascribe to nonverbal animals only crude and rudimentary beliefs. We routinely attribute to a dog beliefs like "The food bowl is empty" and "There is a cat sitting on the fence," but not beliefs like "Either the food bowl is empty or there is no cat sitting on the fence" and "If no cat is sitting on the fence, either it's raining or his master has called him in." It is difficult to think of nonverbal behavior on the basis of which we can attribute to anyone, let alone cats, beliefs with logically complex contents, say, beliefs expressed by "Every cat can be fooled some of the time, but no cat can be fooled all of the time," or "Since tomorrow is Monday, my master will head for work in Manhattan as usual, unless his cold gets worse and he decides to call in sick," and the like. It is arguable that in order to have beliefs or entertain thoughts like these, you must be a language user with a capacity to generate and understand sentences with complex structure.

Confining our attention to language speakers, then, let us see how well the proposed definition of belief works as a behaviorist definition. Difficulties immediately come to mind. First, as we saw with Hempel's "toothache" example, the definition presupposes that the person in question *understands* the question "Is it the case that *p*?"—and understands it *as a request for* an answer of a certain kind. (The definition as stated presupposes that the subject understands English, but this feature of the definition can be eliminated by modifying the antecedent, thus: "S is asked a question in a language S understands that is synonymous with the English sentence 'Is it the case that *p*?'") But understanding is a psychological concept, and if this is so, the proposed definition cannot be considered behavioristically acceptable (unless we have a prior behavioral definition of "understanding" a language). The same point applies to the consequent of the definition: In uttering the words "Yes, it is the case that *p*," S must *understand what these words mean* and *intend them to be understood by her hearer to have that meaning*. It is clear that speech acts like saying something and uttering words with an intention to communicate carry substantial psychological presuppositions about the subject. If they are to count as "behavior," it would seem that they must be classified as type (iii) or (iv) behavior, not as motions and noises.

A second difficulty (this too was noted in connection with Hempel's example): When S is asked the question "Is it the case that *p*?" S responds in the desired way only if S *wants* to tell the truth. Thus, the condition "if S wants to tell the truth" must be added to the antecedent of the definition, but this again threatens its behavioral character. The belief that *p* leads to an utterance of a sentence expressing *p* only if we combine the belief with a certain desire, the desire to tell the truth. The point can be generalized: Often behavior or action issues from a complex of mental states, not from a single, isolated mental state. As a rule, beliefs alone do not produce any specific behavior unless they are combined with appropriate desires.[12] Nor will desires: If you want to eat a ham sandwich, this will lead to your ham-sandwich-eating behavior only if you believe that what you are handed is a ham sandwich; if you believe that it is a beef-tongue sandwich, you may very well pass it up. If this is so, it seems not possible to define belief in behavioral terms without building desire into the definition, and if we try to define desire behaviorally, we find that that is not possible unless we build belief into *its* definition.[13] This would indeed be a very small definitional circle.

The complexity of the relationship between mental states and behavior can be appreciated in a more general setting. Consider the following schema relating desire, belief, and action:

> *Desire-Belief-Action Principle (DBA)*. If a person desires that *p* and believes that doing A is an optimal way to secure that *p*, she will do A.

There are various ways of sharpening this principle: For example, it is probably more accurate to say, "She will try to do A" or "She will be disposed to do A," rather than "She will do A." In any event, some such principle as DBA underlies our "practical reasoning"—the means-ends reasoning that issues in action. It is by appeal to such a principle that we "rationalize" actions—that is, give reasons that explain why people do what they do. DBA is also useful as a predictive tool: When we know that a person has a certain desire and that she takes a certain action as an effective way of securing what she desires, we can reasonably predict that she will do, or try to do, the required action. Something like DBA is often thought to be fundamental to the very concept of "rational action."

Consider now an instance of DBA:

1. If Mary desires that fresh air be let into the room and believes that opening the window is a good way to make that happen, she will open the window.

Is (1) true? If Mary does open the window, we could explain her behavior by appealing to her desire and belief as specified in (1). But it is clear that she may have the desire and belief but not open the window—not if, for example, she thinks that opening the window will also let in the horrible street noise that she abhors. So perhaps we could say:

2. If Mary desires fresh air to be let in and believes that opening the window is a good way to make that happen, but if she also believes that opening the window will let in the horrible street noise, she will not open the window.

But can we count on (2) to be true? Even given the three antecedents of (2), Mary will still open the window if she also believes that her ill mother very badly needs fresh air. It is clear that this process could go on indefinitely.

This suggests something interesting and very important about the relationship between mental states and behavior, which can be stated like this:

Defeasibility of Mental-Behavioral Entailments. If there is a plausible entailment of behavior B by mental states $M_1, \ldots, M_n$, there always is a further mental state M_{n+1} such that $M_1, \ldots, M_n, M_{n+1}$ together plausibly entail not-B.

If we assume not-B (that is, the failure to produce behavior B) to be behavior as well, the principle can be iteratively applied, without end, as we saw with Mary and the window opening: There exists some mental state M_{n+2} such that $M_1, \ldots, M_n, M_{n+1}, M_{n+2}$ together plausibly entail B. And so on without end.[14]

This shows that the relationship between mental states and behavior is highly complex: The moral is that mind-to-behavior connections are always *defeasible*—and defeasible by the occurrence of a *further mental state*, not merely by physical barriers and hindrances (as when Mary cannot open the window because her arms are paralyzed or the window is nailed shut). This makes the prospect of producing for each mental expression a purely behavioral-physical definition extremely remote. But we should not lose sight of the important fact that the defeasibility thesis does state an important and interesting connection between mental phenomena and behavior. The thesis does not say that there are no mental-behavioral entailments—it only says that such entailments are more complex than they might first appear, in that they always face potential mental defeaters.

Let us now turn to another issue. Suppose you want to greet someone. What behavior is entailed by this want? As we might say, greeting desires issue in greeting behavior. But what is greeting behavior? When you see Mary across the street and want to greet her, you might wave to her, cry out "Hi, Mary!" The entailment is defeasible since you would not greet her, even though you want to, if you also thought that by doing so you might cause her embarrassment. Be that as it may, saying that wanting to greet someone issues in a *greeting* does not say much about the *observable physical behavior*, because greeting is an action that includes a manifest psychological component (behavior of type [iii] distinguished earlier). Greeting Mary involves *noticing* and *recognizing* her, *believing* (or *hoping*) that she will *notice* your physical gesture and *recognize* it as expressing your *intention* to greet her, and so on. Greeting obviously will not count as behavior of kind (i) or (ii)—that is, a physiological response or bodily movement.

But does wanting to greet entail any bodily movements? If so, what bodily movements? There are innumerable ways of greeting: You can greet by waving your right hand, waving your left hand, or waving both; by saying "Hi!" or "How are you?" or "Hey, how're you doing, Mary?"; by saying these things in French or Chinese (Mary is from France, and you and Mary are taking a Chinese class); by rushing up to Mary and shaking her hand or giving her a hug; and countless other ways. In fact, any physical gesture will do as long as it is socially recognized as a way of greeting.[15]

And there is a flip side to this. As travel guidebooks routinely warn us, a gesture that is recognized as friendly and respectful in one culture may be taken as expressing scorn and disdain in another. Indeed, within our own culture the very same physical gesture could count as greeting someone, indicating your presence in a roll call, bidding at an auction, signaling for a left turn, and any number of other things. The factors that determine exactly what it is that you are doing when you produce a physical gesture include the customs, habits, and conventions that are in force as well as the particular circumstances at the time—a complex network of entrenched customs and practices, the agent's beliefs and intentions, her social relationships to other agents involved, and numerous other factors.

Considerations like these make it seem exceedingly unlikely that anyone could ever produce correct behavioral definitions of mental terms linking every mental expression with an equivalent behavioral expression referring solely to pure physical behavior ("motions and noises"). In fact, we have seen here how futile it would be to look for interesting generalizations, much less definitions, connecting mental states, like wanting to greet someone, with

physical behavior. To have even a glimmer of success, we would need, it seems, to work at the level of intentional action, not of physical behavior—that is, at the level of actions like greeting a friend, buying and selling, and reading the morning paper, not behavior at the level of motions and noises.

Do Pains Entail Pain Behavior?

Nevertheless, as noted earlier, some mental phenomena seem more closely tied to physical behavior—occurrences like pains and itches that have "natural expressions" in behavior. When you experience pain, you wince and groan and try to get away from the source of the pain; when you itch, you scratch. This perhaps is what gives substance to the talk of "pain behavior"; it is probably easier to recognize pain behavior than, say, greeting behavior, in an alien culture. We sometimes try to hide our pains and may successfully suppress winces and groans; nonetheless, pains do seem, under normal conditions, to manifest themselves in a roughly identifiable range of physical behavior. Does this mean that pains entail certain specific types of physical behavior?

Let us first get clear about what "entailment" is to mean in a context of this kind. When we say that pain "entails" winces and groans, we are saying that "Anyone in pain winces and groans" is *analytically*, or *conceptually*, *true*—that is, like "Bachelors are unmarried" and "Vixens are females," it is true *solely in virtue of the meanings of the terms involved* (or *the concepts expressed by these terms*). If "toothache" is definable, as Hempel claims, in terms of "weeping" and "making gesture G" (where we leave it to Hempel to specify G), toothache entails weeping and making gesture G in our sense. And if pain entails winces and groans, no organism could count as "being in pain" unless it could evince wincing and groaning behavior. That is, there is no "possible world" in which something is in pain but does not wince and groan.[16]

Some philosophers have argued that there is no pain-behavior entailment because pain behavior can be completely and thoroughly suppressed by some people, the "super-Stoics" and "super-Spartans" who have trained themselves not to show their pains in overt behavior.[17] This objection can be met, at least partially, by pointing out that super-Spartans, although they do not actually exhibit pain behavior, can still be said to have a *propensity*, or *disposition*, to exhibit pain behavior—that is, they *would* exhibit overt pain behavior *if certain conditions were to obtain* (for example, the super-Spartan code of conduct is renounced, their inhibition is loosened by alcohol, etc.). It is only that these conditions do not obtain for them, and so their behavior dispositions

associated with pain remain unmanifested. And a truthful super-Spartan will say yes when asked "Are you in pain?" although she will not groan, wince, or complain. There is this difference, then, between a super-Spartan who is in pain and another super-Spartan who is not in pain: It is true of the former, but not the latter, that if certain conditions were to obtain for her, she would exhibit pain behavior. It seems, therefore, that the objection based on the conceivability of super-Spartans can be substantially mitigated by formulating the entailment claim in terms of behavior dispositions or propensities rather than actual behavior production. After all, most behaviorists identify mentality with behavior dispositions, not actual behaviors.[18]

So the modified entailment thesis says this: It is an analytic, conceptual truth that anyone in pain has a propensity to wince or groan. Is this true? Consider animals: Dogs and cats can surely feel pain. Do they wince or groan? Perhaps. How about squirrels or bats? How about snakes and octopuses? Evidently, in order to groan or wince or emit a specified type of behavior (such as screaming and writhing in pain), an organism needs a certain sort of body and bodily organs with specific capacities and powers. Only animals with vocal cords can groan or scream; we can be certain that no one has ever observed a groaning snake or octopus! Thus, the entailment thesis under consideration has the consequence that organisms without vocal cords cannot be in pain, which is absurd. The point can be generalized: Whatever behavior type is picked, we can coherently imagine a pain-capable organism that is physically unsuited to produce behavior of that type.[19]

If this is the case, there is no specific behavior type that is entailed by pain. More generally, the same line of consideration should show that no specific behavior type is entailed by any mental state. And yet a weaker thesis, perhaps something like the following, may be true:

> *Weak Behavior Entailment Thesis*. For any pain-capable species[20] there is a certain behavior type B such that, for that species, being in pain entails a propensity to emit behavior of type B.

According to this thesis, then, each species may have its own special way of expressing pain behaviorally, although there are no universal and species-independent pain-to-behavior entailments. If this is correct, the concept of pain involves the concept of behavior only in this sense: Any organism in pain has a propensity to behave in some characteristic way. Note that the Weak Entailment Thesis is formulated in terms of a "propensity" to exhibit a type of behavior; having a propensity should be taken to mean that only when an appropriate

set of conditions obtains, the phenomenon will occur, or, alternatively, that there is a fairly high probability of its occurring. In any case, it is clear that there is no behavior pattern that can count as "pain behavior" across all pain-capable organisms (and perhaps also inorganic systems). Again, this makes the prospect of defining pain in terms of behavior exceedingly remote.

Ontological Behaviorism

Logical behaviorism is a thesis about the meanings of psychological expressions; as you recall, the claim is that the meaning of every psychological term is definable exclusively on the basis of behavioral-physical terms. More concretely, the claim is that given any sentence including psychological expressions, we can in principle produce a synonymous sentence devoid of psychological expressions. But we can also consider a behaviorist thesis about psychological states or phenomena as such, independently of the language in which they are described. The question—the "ontological" question—is what mental states *are*. Given that psychological sentences are translatable into behavioral sentences, does this mean that there are only behaviors, but no mental states? No pains; only pain behavior?

A radical behaviorist may claim that there are no mental facts over and above actual and possible behavioral facts, and that inner mental events do not exist, and that if they did, they are of no consequence. This is ontological behaviorism: Existentially, our mentality consists solely in behaviors and behavioral dispositions; there is nothing more. This, therefore, is a form of psychological eliminativism,[21] the view that mentality as ordinarily conceived is as misguided and defunct as the phlogiston theory of combustion and the neo-vitalist theory of entelechies as the "principle of life." Like such discredited scientific theory, mentalistic psychology will be jettisoned sooner and later. Such is the claim of radical behaviorism.

Compare the following two claims about pain:

1. Pain = winces and groans.
2. Pain = the cause of winces and groans.

Claim (1) expresses an ontological behaviorism about pain; it tells us what pain is—it is winces and groans. There is nothing more to pain than pain behavior—if there is also some private event going on, that is not pain, or part of pain, whatever it is, and it is psychologically irrelevant. But (2) is not a form of ontological behaviorism, since the *cause* of winces and groans need not be,

and probably isn't, more behavior. Clearly (2) may be affirmed by someone who thinks that it is an *internal state* of organisms (say, a neural state) that causes pain behavior, like winces and groans. Moreover, a dualist—even a Cartesian dualist—can welcome (2): She would say that a private mental event, an inner pain experience, is the cause of winces, groans, and other pain behavior. Further, we might even claim that (2) is analytically or conceptually true: The concept of pain is that of an internal state apt to cause characteristic pain behaviors like winces and groans.[22] Note this paradoxical result: (2) can be taken as an unexpected vindication of logical behaviorism about "pain," since it allows us to translate any sentence of the form "X is in pain" into "X is in a state that causes winces and groans," a sentence devoid of psychological expressions.[23] The same goes for other sentences including the term "pain." On the ontological question "What is pain really?" (2) is consistent with physicalism, property dualism, and even Cartesian interactionist dualism, though not with epiphenomenalism or Leibniz's preestablished harmony. One lesson of all this is that logical behaviorism does not entail ontological behaviorism.

Does ontological behaviorism entail logical behaviorism? Again, the answer has to be no. From the fact that Xs are Ys, nothing interesting follows about the meanings of the expressions "X" and "Y"—in particular, nothing follows about their interdefinability. Consider some examples: We know that bolts of lightning are electric discharges in the atmosphere and that genes are DNA molecules. But the expressions "lightning" and "electric discharge in the atmosphere" are not conceptually related, much less synonymous; nor are the expressions "gene" and "DNA molecule." So it may be that pain = winces, groans, and avoidance behavior. But you would not be able to verify that "pain" means the same as "winces, groans, and avoidance behavior" by consulting the most comprehensive dictionary.

In a similar vein, one could say, as some philosophers have argued,[24] that there are in this world no inner private episodes like pains, itches, and twinges but only observable behaviors or dispositions to exhibit such behaviors. One may say this because one holds a certain form of logical behaviorism or takes a dim view of supposedly private and subjective episodes in an inner theater. But one may be an ontological behaviorist on a methodological ground, affirming that there is *no need to posit* private inner events like pains and itches since they are not needed—nor are they able—to explain observed behaviors of humans and other organisms, as neural-physical states are sufficient for this purpose. A person holding such a view may well concede that the phenomenon purportedly designated by "pain" is an inner subjective state but will insist that there is no reason to think that the word actually refers to anything real (compare with

"witch," or "Bigfoot"). Daniel Dennett has urged that our concept of a private qualitative state ("qualia") is saddled with conditions that cannot be simultaneously satisfied, and that, as a result, there can be nothing that corresponds to the traditional idea of a private inner episode.[25] Paul Churchland and Stephen Stich have argued that beliefs, desires, and other intentional states as conceived in "folk" psychology will go the way of phlogiston and entelechies as systematic, scientific psychology makes progress.[26]

The Real Relationship Between Pain and Pain Behavior

Our discussion has revealed serious difficulties with any entailment claims about the relationship between pain and pain behavior, or more generally, between types of mental states and types of behavior. The considerations seemed to show that though our pains may cause our pain behaviors, this causal relation is a contingent fact. But leaving the matter there is unsatisfying: Surely pain behaviors—groans, winces, screams, writhings, attempts to get away, and such—have something important to do with our notion of pain. How else could we learn, and teach, the concept of pain or the meaning of the word "pain"? Wouldn't we rightly deny the concept of pain to a person who does not at all appreciate the connection between pains and these characteristic pain behaviors? If a person observes someone writhing on the floor, clutching his broken leg, and screaming for help and yet refuses to acknowledge that he is in pain, wouldn't it be correct to say that this person does not have the concept of pain, that he does not know what "pain" means? If Wittgenstein's "beetle in the box" shows anything, it is the point that publicly accessible behaviors are essential to anchor the meanings of our mental terms, like "pain," and explain the possibility of knowledge of what goes on in other people's minds. That is, observable behavior seems to have an essential grounding role for the semantics of our psychological language and the epistemology of other minds. What we need, therefore, is a positive account of the relationship between pain and pain behavior that explains their intimate connection without making it into one of logical, or conceptual, entailment.

The following is one possible story. Let us begin with an analogy: How do we fix the meaning of "one meter long"—that is, the concept of a meter? We sketch an answer based on Saul Kripke's influential work on names and their references.[27] Consider the Standard Meter: a bar of platinum-iridium alloy kept in a vault near Paris.[28] Is the following statement necessarily, or analytically, true?

The Standard Meter is one meter long.

There is a clear sense in which the Standard Meter *defines* what it is to be one meter in length. But does being the Standard Meter (or having the same length as the Standard Meter) entail being one meter long? The Standard Meter is a particular physical object, manufactured at a particular date and place and now located somewhere in France, and surely this metallic object might not have been the Standard Meter and might not have been one meter long. (It could have been fashioned into a bowl, or it could have been made into a longer rod of two meters.) In other words, it is a contingent fact that this particular platinum-iridium rod was selected as the Standard Meter, and it is a contingent fact that it is one meter long. No middle-sized physical object has the length it has necessarily; anything could be longer or shorter—or so it seems. We must conclude, then, that the statement that something has the same length as the Standard Meter does not logically entail that it is one meter long, and it is not analytically, or conceptually, true that if the length of an object coincides with that of the Standard Meter, it is one meter long.

But what, then, is the relationship between the Standard Meter and the concept of the meter? After all, the Standard Meter is not called that for nothing; there must be some intimate connection between the two. A plausible answer is that we specify the property of being one meter long (the meaning, if you wish, of the expression "one meter long") by the use of a *contingent* relationship in which the property stands. One meter is the length of this thing (namely, the Standard Meter) here and now. It is only contingently one meter long, but that is no barrier to using it to specify what counts as one meter. This is just like when we point to a ripe tomato and say, "Red is the color of this tomato." It is only a contingent fact that this tomato is red (it could have been green), but we can use this contingent fact to specify what the color red is and what the word "red" means.

Let us see how a similar account might go for pain: We specify what pain is (or fix the meaning of "pain") by reference to a contingent fact about pain, namely, that pain causes winces and groans in humans. This is a contingent fact about this world. In worlds in which different laws hold, or worlds in which the central nervous systems of humans and those of other organisms are hooked up differently to peripheral sensory surfaces and motor output systems, the patterns of causal relations involving pain may be very different. But as things stand in this world, pain is the cause of winces and groans and certain other behaviors in humans and related animal species. In worlds in which pains do not cause winces and groans, different behaviors may count

as pain behavior, in which case pain specifications in those worlds could advert to the behaviors caused by pains there. This is similar to the color case: If cucumbers but not ripe tomatoes were red, we would be specifying what "red" means by pointing to cucumbers instead.

The foregoing is only a sketch of an account but not an implausible one. It explains how (2) above ("Pain = the cause of winces and groans"), though only contingently true, can help specify what pain is and fix the reference of the term "pain." And it seems to show a good fit with the way we learn, and teach, how to use the word "pain" and other mental expressions denoting sensations. The approach brings mental expressions under the same rubric with many other expressions, as we have seen, such as "red" and "one meter long." Though not implausible, the story may not be over just yet: The reader is encouraged to think about the possible differences between the case of pain and cases like the color red and one meter in length. In particular, think about how it deals with, or fails to deal with, the conundrum of Wittgenstein's "beetle in the box."

Behaviorism in Psychology

So far we have been discussing behaviorism as a philosophical doctrine concerning the meanings of mental terms and the nature of mental states. But as we noted at the outset, "behaviorism" is also the name of an important and influential psychological movement initiated early in the twentieth century that came to dominate scientific psychology and the social sciences in North America and many other parts of the world for several decades. It held its position as the reigning methodology of the "behavioral sciences" until the latter half of the century, when "cognitivism" and "mentalism" began a strong comeback and replaced it as the new orthodoxy.

Behaviorism in science can be viewed in two ways: First, as a precept on how psychology should be conducted as a science, it provides guidance to questions like what its proper domain should be, what conditions should be placed on admissible evidence, what its theories are supposed to accomplish, by what standards its explanations are to be evaluated, and so on. Second, behaviorism, especially B. F. Skinner's "radical behaviorism," is a specific behaviorist research paradigm seeking to construct psychological theories conforming to a fairly explicit and precisely formulated pattern (for example, Skinner's "operant conditioning"). Here we have room only for a brief and sketchy discussion of scientific behaviorism in the first sense. Discussion of Skinner's radical behaviorism is beyond the scope of this book.

We can begin with what may be called methodological behaviorism:

(I) The only admissible evidence for the science of psychology is observable behavioral data—that is, data concerning the observable physical behavior of organisms.

We can understand (I) somewhat more broadly than merely as a stricture on admissible "evidence" by focusing on the "data" it refers to. Data serve two closely related purposes in science: First, they constitute the domain of phenomena for which theories are constructed to provide explanations and predictions; second, they serve as the evidential basis that can support or undermine theories. What (I) says, therefore, is that psychological theories should attempt to explain and predict only data concerning observable behavior and that only such data should be used as evidence against which psychological theories are to be evaluated. These two points can be seen to collapse into one when we realize that explanatory and predictive successes and failures constitute, by and large, the only measure by which we evaluate how well theories are supported by evidence.

The main reason some psychologists and philosophers have insisted on the observability of psychological data is to ensure the *objective* or *intersubjective testability* of psychological theories. It is thought that introspective data—data obtained by a subject by inwardly inspecting her own inner Cartesian theater—are essentially private and subjective and hence cannot serve as the basis for intersubjective validation of psychological theories. In short, the idea is that intersubjective access to data is required to ensure the possibility of intersubjective agreement in science and that the possibility of intersubjective agreement is required to ensure the objectivity of psychology. Only behavioral (and more broadly, physical) data, it is thought, meet the condition of intersubjective observability. In short, (I) aims at securing the objectivity of psychology as a science.

What about a subject's verbal reports of her inner experiences? A subject in an experiment involving mental imagery might report: "I am now rotating the figure counterclockwise." What is wrong with taking the following as an item of our data: Subject S is rotating her mental image counterclockwise? Someone who holds (I) will say something like this: Strictly speaking, what we can properly consider an item of data here is S's utterance of the words "I am now rotating the figure counterclockwise." Counting S's actual mental operation of rotating her mental image as a datum involves the assumption that she is a

competent speaker of English, that intersubjective meaning can be attached to reports of inner experience, and that she is reporting her experience correctly. These are all substantial psychological assumptions, and we cannot consider the subject's reports of her visual activity to meet the criterion of intersubjective verifiability. Therefore, unless these assumptions themselves can be behaviorally justified, the cognitive scientist is entitled only to the subject's utterance of the string of words, not the presumed content of those words, as part of her basic data.

Consciousness is usually thought to fall outside the province of psychological explanation for the behaviorist. Inner conscious states are not among the phenomena it is the business of psychological theory to explain or predict. In any case, many psychologists and cognitive scientists may find (I) by and large acceptable, although they are likely to disagree about just what is to count as *observable* behavior. (Some may consider verbal reports, with their associated meanings, as admissible data, especially when they are corroborated by nonverbal behavior.)

A real disagreement arises, though, concerning the following stronger version of methodological behaviorism:

> (II) Psychological *theories* must not invoke the *internal states* of psychological subjects; that is, psychological explanations must not appeal to internal states of organisms, nor should references to such states occur in deriving predictions about behavior.

This appears to have been a tenet of Skinner's psychological program. On this principle, organisms are to be construed as veritable black boxes whose internal structure is forever closed to the psychological investigator. Psychological generalizations, therefore, must only correlate observable stimulus conditions as input, behavioral outputs, and subsequent reinforcements. But isn't it obvious that when the same stimulus is applied to two organisms, they can respond with different behavior output? How can we explain behavioral differences elicited by the same stimulus condition without invoking differences in their internal states?

The Skinnerian answer is that such behavioral differences can be explained by reference to the differences in the *history* of reinforcement for the two organisms; that is to say, the two organisms emit different behavior in response to the same stimulus because their *histories* involving external stimuli, elicited behaviors, and the reinforcements following the behaviors are different. But if such an explanation works, isn't that because the differences in the

histories of the two organisms led to differences in their present internal states? Isn't it plausible to suppose that these differences *here and now* are what is directly implicated in the production of different behaviors *now*? To suppose otherwise would be to embrace "mnemic" causation—causal influence that leaps over a temporal gap with no intermediate links bridging cause and effect. Apart from such metaphysical doubts, there appears to be an overwhelming consensus at this point that the stimulus-response-reinforcement model is simply inadequate to generate explanatory or predictive theories for vast areas of human and animal behaviors.

And why is it impermissible to invoke present internal differences as well as differences in histories to explain differences in behavior output? Notice how sweeping the constraint expressed by (II) really is: It outlaws references not only to inner mental states of the subject but also to its internal physical-biological states. Methodological concerns with the objectivity of psychology as a science provide an intelligible (if perhaps not sufficient) motivation for banishing the former, but it seems clearly insufficient to justify banning the latter from psychological theories and explanations. Even if it is true, as Skinner claims,[29] that invoking internal neurobiological states does not help psychological theorizing, that hardly constitutes a sufficient ground for prohibiting it as a matter of scientific methodology.

In view of this, we may consider a further version of behaviorism as a rule of psychological methodology:

> (III) Psychological theories must make no reference to inner *mental* states in formulating psychological explanations.

This principle allows the introduction of internal biological-physical states, including states of the central nervous system, into psychological theories and explanations, prohibiting only reference to inner mental states. But what is to count as such a state? Does this principle permit the use of such concepts as "drive," "information," "memory," "attention," "mental representation," and the like in psychological theories? To answer this question we would have to examine these concepts in the context of particular psychological theories making use of them; this is not a task for armchair philosophical conceptual analysis. We should keep in mind, though, that the chief rationale for (III)—in fact, the driving motivation for the entire behaviorist methodology—is the insistence on the objective testability of theories and public access to sharable data. This means that what (III) is intended to prohibit is the introduction of *private subjective* states for which objective access is thought to be problematic, not the use

of theoretical constructs posited by psychological theories for explanatory and predictive purposes, as long as these meet the requirement of intersubjectivity. Unlike overt behavior, these constructs are not, as a rule, "directly observable," and they are not strictly definable or otherwise reducible in terms of observable behavior. However, they differ from the paradigmatic inner mental states in that they apparently do not show the first-person/third-person asymmetry of epistemic access. Scientific theories often introduce theoretical concepts for entities (electrons, magnetic fields, quarks) and properties (spin, polarization) that go far beyond the limits of human observation. Like any other science, psychological theory should be entitled to such theoretical constructs.

But in excluding private conscious states from psychological theory, (III) excludes them from playing any causal-explanatory role in relation to behavior. If it is true, as we ordinarily think, that some of our behavior is caused by inner mental states disallowed by (III), our psychological theory is likely to be incomplete: There may well be behavior for which no theory meeting (III) can provide full explanations. (Some of these issues are discussed further in chapters 7, 9, and 10.)

Are there other methodological constraints for psychological theory? How can we be sure that the states and entities posited by a psychological theory (for example, "intelligence," "mental representation," "drive reduction") are "real"? If, in explaining the same data, one psychological theory posits one set of unobservable states and another theory posits an entirely different set, which theory, if any, should be believed? That is, which theory represents the *psychological reality* of the subjects? Does it make sense to raise such questions? If it does, should there be the further requirement that the entities and states posited by a psychological theory have a "biological reality"—that is, must they somehow be "realized" or "implemented" in the biological-physical structures and processes of the organism? These are important questions about the science of psychology, and we deal with some of them later in our discussion of mind-body theories and the status of cognitive science (chapters 5, 6).

Why Behavior Matters to Mind

Our discussion thus far has been, by and large, negative toward behaviorism. This should not be taken to mean that we should take a negative attitude on the relevance of behavior to minds. The fact is that the importance of behavior to mentality cannot be overemphasized. In retrospect, it seems that, impressed by the crucial role of behavior to mentality, various forms of behaviorism, in

particular logical behaviorism, got carried away, going way overboard and advocating extreme and unrealistic theories, with a reformer's zeal.

There are three main players on the scene in discussions of mentality: mind, brain, and behavior. An important task of the mind-body problem is to elucidate the relationships among these three elements. The detailed issues and problems are yet to be discussed in the rest of this book. But here is a rough picture:

1. The brain is the ontological—that is, existential—base of the mind.
2. The brain, and perhaps the mind also, is the cause of behavior.
3. Behavior is the semantic foundation of mental language. It is what fixes the meanings of our mental/psychological expressions.
4. Behavior is the primary, almost exclusive, evidence for the attribution of mental states to other beings with minds. Our knowledge of other minds depends primarily on observation of behavior.

It is fair to say these statements are what most of us believe. There will be dissenters, especially about (1) and (2)—for example, Cartesian dualists. What concerns us here are items (3) and (4). Without behavior, it is hard to see how our mental terms can acquire their common, public meanings fit for interpersonal communication. And without behavioral evidence (including verbal behavior), it is not possible to know what others are thinking and feeling. (Try to imagine how you might find out what an immaterial soul is thinking or feeling.) If we were to lose observational access to others' behavior, the fabric of our social relationships would completely unravel. Unquestionably, behavior is the semantic and epistemological foundation of our mental and social life.

To summarize, the brain is what existentially underlies, and supports, our mental life. You take away the brain, and mental life is no more. Behavior, on the other hand, is the semantical and epistemological foundation of mentality. Without it, psychological language would be impossible, and we could never know what goes in other minds. It is impossible to exaggerate, or even underplay, the crucial place observable behavior has in our social life.

For Further Reading

The influential classic work representing logical behaviorism is Gilbert Ryle, *The Concept of Mind*. Also important are Rudolf Carnap, "Psychology in Physical Language," and Carl G. Hempel, "The Logical Analysis of Psychology."

For an accessible Wittgensteinian perspective on mind and behavior, see Norman Malcolm's contributions in *Consciousness and Causality* by D. M. Armstrong and Norman Malcolm. For scientific behaviorism, see B. F. Skinner's *Science and Human Behavior* and *About Behaviorism*. Both are intended for nonspecialists.

For a historically important critique of Skinnerian behaviorism, see Noam Chomsky's review of Skinner's *Verbal Behavior*. For criticism of logical behaviorism, see Roderick M. Chisholm, *Perceiving*, pp. 173–185; and Hilary Putnam, "Brains and Behavior." George Graham's article "Behaviorism" in the *Stanford Encyclopedia of Philosophy* is a useful resource; so is Georges Rey's entry "Behaviorism" in the *Macmillan Encyclopedia of Philosophy*, 2nd ed.

Notes

1. William James, *The Principles of Psychology*, p. 15. Page references are to the 1981 edition.

2. Ibid., p. 185.

3. J. B. Watson, "Psychology as the Behaviorist Views It," p. 158.

4. William James, *The Principles of Psychology*, p. 21 (emphasis in original).

5. This piquant term comes from Daniel C. Dennett, *Consciousness Explained*. Dennett considers the Cartesian theater an incoherent myth.

6. Ludwig Wittgenstein, *Philosophical Investigations*, section 293. We need to assume that there are no beetles flying around for everyone to see!

7. For the notion of behavior as internally caused bodily motion, see Fred Dretske, *Explaining Behavior*, chapters 1 and 2.

8. You might feel uncomfortable about the last two examples: Perhaps our bodies do these things, but it sounds odd to say that *we* do these things. The ordinary notion of doing seems to involve the idea of voluntariness; however, the notion of behavior appropriate to behaviorism need not include such an element.

9. Carl G. Hempel, "The Logical Analysis of Psychology," p. 91.

10. Positivists, including Hempel, often used a much looser sense of definition (and translatability); however, for logical behaviorism to be a significant thesis, we need to construe definition in a more strict sense.

11. Carl Hempel, "The Logical Analysis of Psychology," p. 17.

12. There is a long-standing controversy in moral theory as to whether certain beliefs (for example, the belief that you have a moral duty to help a friend), without any associated desires, can motivate a person to act. The dispute, however, concerns only a small class of beliefs, chiefly evaluative and

normative beliefs about what ought to be done, what is desirable, and the like. The view that to generate an action both desire and belief must be present is usually attributed to Hume.

13. For an early statement of this point, see Roderick M. Chisholm, *Perceiving*.

14. This may be what is distinctive and interesting about the "ceteris paribus" clauses qualifying psychological generalizations—in particular, those concerning motivation and action.

15. The phenomena discussed in this paragraph and the next are noted in Berent Enç, "Redundancy, Degeneracy, and Deviance in Action."

16. Strictly speaking, this last sentence defines "metaphysical" entailment, as distinguished from analytical or conceptual entailment as defined earlier. There are differences between them that can be important in some context; however, this will not affect our discussion.

17. Hilary Putnam, "Brains and Behavior."

18. Moreover, many mental states have bodily manifestations; pain may be accompanied by a rise in blood pressure and a quickening pulse, and super-Spartans presumably could not "hide" these physiological signs of pain (recall Hempel's behavioral translation of "Paul has a toothache"). Whether these count as "behavior" may only be a verbal issue in this context.

19. Perhaps this can be called a "multiple realizability" thesis in regard to behavior. On multiple realizability, see chapter 5. Whether it has consequences for behaviorism that are similar to the supposed consequences of the multiple realizability of mental states is an interesting further question.

20. Species may be too wide here, given that expressions of pain are, at least to some extent, culture-specific and can even differ from person to person within the same culture.

21. See Paul Churchland, "Eliminative Materialism and the Propositional Attitudes."

22. See chapters 5 and 6 for discussion of the functionalist conception of pain as that of a "causal intermediary" between certain stimulus conditions (for example, tissue damage) and characteristic pain behaviors.

23. This does not mean that the original logical behaviorists, like Hempel and Gilbert Ryle, would have accepted (2) as a behavioral characterization of "pain." The point, however, is that it meets Hempel's translatability thesis—his form of logical behaviorism. Note that the "cause" is a topic-neutral term—it is neither mental nor behavioral-physical.

24. See Gilbert Ryle, *The Concept of Mind*.

25. Daniel Dennett, "Quining Qualia."

26. Paul Churchland, "Eliminative Materialism and the Propositional Attitudes"; Stephen Stich, *From Folk Psychology to Cognitive Science: The Case Against Belief*.

27. See Saul Kripke, *Naming and Necessity*.

28. The meter is no longer defined this way; the current definition, adopted in 1984 by the General Conference on Weights and Measures, is reportedly based on the distance traveled by light through a vacuum in a certain (very small) fraction of a second.

29. See B. F. Skinner, *Science and Human Behavior*.