

“Mental States are like Diseases”

Behaviorism in the Immanuel Kant Lectures

Sander Verhaegh*

0. Abstract

One of the great values of the Immanuel Kant Lectures is that it sheds new light on the nature of Quine’s views about behaviorism. Where Quine’s linguistic behaviorism is well-known, the Lectures contain one of his most detailed discussions of behaviorism in psychology and the philosophy of mind. Quine clarifies the nature of his psychological commitments by arguing for a view that is quite modest: he argues against ‘excessively restrictive’ variants of behaviorism while maintaining that ‘a good measure of behaviorist discipline is still needed to keep [our mental] terms under control’. In this paper, I use Quine’s comments in the Lectures to reconstruct his position. I start by distinguishing three types of behaviorism in psychology and the philosophy of mind: ontological behaviorism, logical behaviorism, and epistemological behaviorism. Next, I reconstruct Quine’s perspective on each of these views and argue that he does not fully accept any of them. Finally, I combine these perspectives and reconstruct Quine’s subtle view about behaviorism in psychology.

* Tilburg Center for Logic, Ethics, and Philosophy of Science (TiLPS), Tilburg University, Dante Building (second floor), Warandelaan 2, 5037AB Tilburg, The Netherlands. Email: A.A.Verhaegh@uvt.nl. Website: <http://www.sanderverhaegh.com>

1. Introduction

The nature of Quine's commitment to behaviorism is perhaps best summarized by his slogan: "In psychology one may or may not be a behaviorist, but in linguistics one has no choice" (1987, 341; 1990, 37-8). Quine's *linguistic* behaviorism is well known; throughout his career, he has extensively argued that language is a social art—that we can only acquire a language "by observing other people's verbal behavior" and by having our "own faltering verbal behavior observed and reinforced or corrected by others" (ibid.). Quine's linguistic behaviorism, in other words, tells us something about the data we have available in learning a language; 'mental entities', even if they exist, play no role in language acquisition.¹

Quine's views about *psychological* behaviorism, on the other hand, seem less well defined. Not only because of his claim that 'one may or may not be a behaviorist' in psychology, but also because he has written remarkably little about issues in psychology and the philosophy of mind. In the first decades of his career, for example, Quine never publicly qualified his commitment to behaviorism beyond his views about language learning. Although it is widely recognized that it is a mistake to believe that Quine's approach to language depends on a behaviorist approach to psychology (see, for example, Hylton 2007, 102), many scholars seem to assume that his views on mind and behavior were at the very least compatible with those of the behavioral psychologists.

One of the great values of the Immanuel Kant Lectures is that it sheds new light on the nature of Quine's views about psychological behaviorism. "Mind and Its

Place in Nature”, the first lecture of the series, contains one of Quine’s most detailed discussions of topics and arguments that are mostly ignored in his published work. Quine clarifies the nature of his psychological commitments by arguing for a view that is quite modest: he argues against “excessively restrictive” variants of behaviorism while maintaining that “a good measure of behaviorist discipline is still needed to keep [our mental] terms under control” (p. 5).

In this paper, I use Quine’s comments in the Lectures to reconstruct his position. I start by distinguishing three types of behaviorism in psychology and the philosophy of mind: ontological behaviorism, logical behaviorism, and epistemological behaviorism (section 2). Next, I reconstruct Quine’s perspective on each of these views and argue that he does not fully accept any of them (sections 3-5). Finally, I combine these perspectives and reconstruct Quine’s surprisingly subtle view about behaviorism in psychology. (section 6).

2. Three types of behaviorism

In order to disentangle Quine’s complex position about behaviorism in psychology, we need to distinguish between three types of behaviorism: ontological behaviorism, logical behaviorism, and epistemological behaviorism. Roughly speaking, these variants all reject psychological theories that appeal to mental entities (including folk-psychological ‘theories’) and replace them with theories that redefine psychology as the study of behavior. The difference between these variants, however, is that they all rely on distinct arguments against mentalism; ontological behaviorists argue that mental entities do not exist, logical behaviorists argue that mental statements are meaningless if they cannot be translated into a physicalistic language, and

epistemological behaviorists maintain that mental states, images, and processes are explanatorily redundant.

Before I characterize these behaviorisms in more detail, it should be noted that my classification here is primarily conceptual; it mainly distinguishes between three arguments against mentalism. As a result, the positions sketched here are not mutually exclusive; it is possible to reject mentalism for more than one reason. Nor is my taxonomy intended to be complete. One branch of behaviorism I will not be discussing, for example, is methodological behaviorism; the view that mental states, even if they exist, cannot be studied because introspection is unreliable.² Furthermore, because there are many ways to spell out the view that psychology is a science of behavior, there are countless variants of the views specified in this section. Indeed, most variants of behaviorism that have actually been defended by twentieth-century psychologists combine one or more of the arguments developed below with a specific positive proposal.³ The primary use of my taxonomy here is to separate between distinct arguments against what Quine calls “uncritical mentalism” (1970, 5) in order to obtain a more structured overview of the *philosophical* presuppositions that guide behaviorists in psychology.

Prima facie, *ontological behaviorism* is the most straightforward variant of behaviorism. To accept the ontological argument is simply to deny that mental entities exist. Although it will be quite difficult to convincingly argue that there are no mental entities without relying on epistemological considerations—ontological arguments are often epistemological arguments in disguise—the view itself seems to be reasonably clear-cut; psychologists cannot appeal to mental states, mental processes, and mental images because there are no such things.

In practice, however, the situation is more complicated. Most psychologists and philosophers who dismiss mentalism on ontological grounds will not flatly deny

that mental entities exist. Rather, they will argue that mental entities are not what they seem to be. When Watson argues that “thought processes *are really* motor habits in the larynx” (1913, 177, my emphasis), for example, he is not denying that we do not think, he is reducing thinking to sensori-motor events; he is reducing mental states and processes to states and processes of a non-mental kind.

Reduction is also the aim of the second type of behaviorism specified above: logical behaviorism. According to logical behaviorists, meaningful psychological statements either do not contain mental concepts or they can be *translated* into statements that do not contain mental concepts. In “The Logical Analysis of Psychology”, for example, Carl Hempel argues that any meaningful statement that contains a mental term can be translated without loss into a conjunction of all the physical test sentences which, if true, would verify that statement:

the meaning of a proposition is established by the conditions of its verification. In particular, two differently formulated propositions have the same meaning or the same effective content when, and only when, they are both true or both false in the same conditions [...]. Let us [...] examine a proposition which involves a psychological concept, for example: “Paul has a toothache.” What is the specific content of this proposition [...]? It will be sufficient to indicate some test sentences which describe these circumstances.

- 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.

[...] The proposition in question, which is about someone's "pain", [...] therefore [...] can be re-translated without loss of content into a proposition which no longer contains the term "pain", but only physical concepts. (Hempel 1935/1949, 377-8)

A statement like "Paul has a toothache", in other words, can be translated into a long conjunction of directly verifiable sentences about his tooth and his pain-related behavior.⁴

Although both logical and ontological behaviorists aim to reduce the mental to the physical, logical behaviorism should not be considered to be a branch of ontological behaviorism. Where ontological behaviorists deny that mental entities exist, logical behaviorists aim to *circumvent* ontological questions by maintaining that statements containing mental concepts are either meaningless or abbreviations of statements that do not contain mental concepts. Or, as Hempel expresses it:

Logical behaviorism claims neither that minds, feelings, inferiority complexes, voluntary actions, etc., do not exist, nor that their existence is in the least doubtful. It insists that the very question as to whether these psychological constructs really exist is already a pseudo-problem, since these notions in their 'legitimate use' appear only as abbreviations in physicalistic statements. (Hempel 1935/1949, 381).

Not all behaviorists worry about the verifiability of psychological statements however. In fact, many behaviorists accept some sort of distinction between the mental and the physical, or, at the very least, a distinction between states and processes that can be publicly observed and states and processes that can only be

observed privately. Where logical behaviorists maintain that all meaningful psychological statements are public, behaviorists like Skinner admit that “a small part of the universe is private” for every individual:

The individual’s response to an inflamed tooth [...] is unlike the response which anyone else can make to that particular tooth, since no one else can establish the same kind of contact with it. Events which take place during emotional excitement or in states of deprivation are often uniquely accessible for the same reason; in this sense our joys, sorrows, loves, and hates are peculiarly our own. (1953: 257)

Instead, Skinner offers an *epistemological* argument against mentalism. According to Skinner, any appeal to mental states, processes, and images is explanatorily redundant: “The objection is not that these things are mental but that they offer no real explanation and stand in the way of a more effective analysis” (1969, 222). To see why, consider again the case of Paul’s inflamed tooth. The claim that Paul is in pain, according to Skinner, does not even begin to explain *why* Paul exhibits the behavior listed by Hempel. Paul’s inner state—or better “inner behavior”—is not an *explanans*, it is itself an *explanandum*:

When an example of maladjusted behavior is explained by saying that the individual is ‘suffering from anxiety,’ we have still to be told the cause of the anxiety. But the external conditions which are then invoked could have been directly related to the maladjusted behavior. Again, when we are told that a man stole a loaf of bread because ‘he was

hungry,’ we have still to learn of the external conditions responsible for the ‘hunger.’ These conditions would have sufficed to explain the theft. The objection to inner states is not that they do not exist, but that they are not relevant in a functional analysis. (Skinner 1953, 35)

Epistemological behaviorism, in other words, simply claims that we do not require mental entities in explaining human and non-human behavior. For both our inner states and our outer behavior require an explanation in terms of external conditions.

3. Radical reduction

I have briefly distinguished three types of behaviorism: logical behaviorism, ontological behaviorism, and epistemological behaviorism. In what follows, I will answer the question whether and to what extent Quine accepts these views.

Prima facie, there are some good reasons to suppose that Quine was committed to logical behaviorism—to the view that all psychological statements can be translated into statements that do not contain mental concepts. After all, one of the strongest arguments for logical behaviorism is the observation that we have all learned how to speak our language ‘by observing other people’s verbal behavior’ and by having our ‘own faltering verbal behavior observed and reinforced or corrected by others’. If language is a social art, as Quine maintains, why not assume that mental concepts are social too? Why not assume that our statements about toothaches can be reduced to sentences about the behavioral facts that we *as a community* rely on in teaching each other to talk about toothaches? Indeed, in some passages of the

Immanuel Kant Lectures, Quine seems to come close to this conclusion:

Introspect our mental states as we will, how do we know what to call them? How did we learn to call our anxieties anxieties, our dull aches dull aches, our joys joys and our awareness awareness? Why do we suppose that what we call joys and anxieties are what other people call by those names? Clearly the answer is that such terms are applied in the light of publicly observable symptoms. (IKL, 2-3)

If psychological statements are “socially inculcated and controlled” (1965, 50), in other words, it seems plausible to conclude that they do not depend on anything that cannot be discovered in publicly observable behavior.

Still, it would be a mistake to view Quine as a logical behaviorist. For Hempel’s translation thesis is an example of ‘radical reductionism’, one of the two positivist commitments Quine rejects in “Two Dogmas of Empiricism” (1951a), the paper in which he famously argues that individual statements do not have any distinct empirical content of their own. Against the view that every scientific hypothesis can be translated into a set of directly testable statements, Quine argues that only clusters of scientific hypotheses are testable—that “our statements about the external world face the tribunal of sense experience not individually but only as a corporate body” (p. 41).

Interestingly, Hempel himself has played an important role in developing and popularizing Quine’s holism—offering a wide range of arguments against the strict translationism he professed in the early 1930s. It is therefore not surprising that Hempel too dismisses logical behaviorism in later stages of his career. Indeed, in a reprint of “The Logical Analysis of Psychology”, Hempel adds the following footnote:

I now (1947) consider the type of physicalism outlined in this paper as too restrictive; the thesis that all statements of empirical science are *translatable*, without loss of theoretical content, into the language of physics, should be replaced by the weaker assertion that all statements of empirical science are *reducible* to sentences in the language of physics, in the sense that for every empirical hypothesis, including, of course, those of psychology, it is possible to formulate certain test conditions in terms of physical concepts which refer to more or less directly observable physical attributes. But those test conditions are not asserted to exhaust the theoretical content of the given hypothesis in all cases. (1935/1949, 373n1)

Instead of maintaining that psychological statements can be *translated* into a behavioral language, in other words, Hempel argues that a mental term can be *partially* defined by specifying the experimental conditions that have to obtain in order to find out whether or not the term applies.⁵

Quine, however, rejects even Hempel's weakened version of logical behaviorism. According to Quine, criteria that aim to reduce individual statements to sets of experimental conditions are still too strong. For many respectable physical concepts (e.g. 'absolute temperature' or ' ψ function') will resist any such interpretation. Where partial definitions aim to specify the necessary conditions for the application of a term, Quine shows that experimental test results will never strictly decide whether or not a certain disposition is present.⁶

Given the connection between logical behaviorism and radical reductionism, we should not be surprised that Quine dismisses Hempel's theses in the Immanuel Kant

Lectures:

If [...] in an access of behaviorism we were to refuse any mental states but what are supported by full behavioral criteria across the board, the physical status of the admitted ones would be secure. [...] The physicalist could settle for the hypothetical physical basis and simply label it with the mentalistic term. However, this extreme of behaviorism would be excessively restrictive [...] We do have to depend on behavioral criteria to legitimize and support the terms that we take over from the mental vocabulary, but that support is a matter of degree. (IKL, 4-5).

Instead of claiming that the acceptability of a mental term depends on the question whether statements containing the term can be reduced to statements that only refer to outward behavior, Quine claims that the acceptability of a mental term depends on the question whether or not adopting the term contributes to our overall theory of the world. Indeed, in a paper entitled “Sellars on Behaviorism, Language, and Meaning” (written in the same year as the Immanuel Kant lectures), Quine makes exactly this point:

Mentalistic predicates can be tolerated in the manner of theoretical predicates of physics, e.g. electron spin, or even electron. For them there is no observational criterion, except as these predicates contribute to the coherence and simplicity of an inclusive theory for which there is observational support as a whole. In a word, we can admit them as hidden variables.⁷ (1980a, 126)

Quine, in sum, is not a logical behaviorist. He rejects both translational and more loosely reductionist versions of the position. We can legitimately adopt a mental term if it contributes to our overall theory of the world; statements containing mental terms do not have to be explained away.

Still, we began our discussion by noting that there seems to be a conceptual connection between logical behaviorism and Quine's thesis that language is a social art. If Quine is not a logical behaviorist, then how does he combine his view about mental terms with his view about language learning? How can he maintain that we can only learn how to use a term like 'toothache' through 'social emulation and social feedback' *and* claim that statements containing the term 'toothache' cannot be reduced to statements about publicly observable behavior?

Again, the Immanuel Kant Lectures provide the solution. Quine shows that although we *initially* learn a term by keying it to publicly observable behavior, we soon learn how to extrapolate these terms to situations that are not intersubjectively accessible:

[Mental] terms are applied in the light of publicly observable channels and then extrapolated along private channels. Someone observes my joyful or anxious expression or perhaps observes my gratifying or threatening situation itself, or hears me tell about it. She then applies the word 'joy' or 'anxiety'. After perhaps another such lesson or two, I find myself applying those words to some of my subsequent states on the strength of a felt similarity. I thus take to reporting my joy or anxiety in cases where no outward signs are to be observed beyond my report itself. (IKL, 3)

Although we initially rely on publicly observable behavior in learning how to use mental predicates, in other words, we eventually go beyond such behavior because we ‘extrapolate along private channels’. At some point, I notice that every time someone describes my behavior as anxious, I also privately experience a certain emotional disturbance; I can then extrapolate by describing myself as anxious in all situations in which I seem to experience a similar disturbance, even if in some of those situations there are no outward signs of my anxiety.

Quine’s theory about mental predicates, in sum, recognizes two forces that pull in opposite directions: as long as mental predicates contribute to our theory as a whole, there is no problem if they are only “loosely anchored” to the publicly observable world. On the other hand, “a good measure of behaviorist discipline is still needed to keep the terms under control” (IKL, 5). Or, as Quine expresses it in an as of yet unpublished speech for psychologists, again written in the same year as the Immanuel Kant Lectures:

The behavioral evidence need not be present on every occasion, but still we are well advised not to make free with mentalistic terms that are habitually beyond the control of behavioral criteria. This is where behaviorist discipline comes in. It is one point of strategy, and there is also a complementary point of strategy. As in any theoretical science, there can sometimes be occasion to posit a theoretical force or other entity that promises to systematize and simplify the hypothetical mechanism; even though this entity be pretty remote from what is observable. This sort of theoretical utility or promise, then, is what we should assess when we

allow a mentalistic term to stray very far from the control of behavioral criteria.⁸ (1980c, item 2999, my transcription)

4. Token physicalism, type dualism

I have shown that Quine is not a logical behaviorist, i.e. that he rejects the claim that mental terms are acceptable only if statements about the mental can be reduced to statements about behavior. This does not imply, however, that Quine believes that we *do* require mental terms in our theorizing. Indeed, considering Quine's claim that mental terms are acceptable if they 'contribute to the coherence and simplicity of an inclusive theory for which there is observational support as a whole', we still need to answer the question whether or not mental terms *in fact* 'systematize and simplify' our theory of the world.

Perhaps it will not be a surprise that Quine answers this question negatively. After all, Quine is a staunch physicalist—he accepts that “nothing happens in the world, not the flutter of an eyelid, not the flicker of a thought, without some redistribution of micro-physical states” (1978b, 98). Indeed, on the very first page of the Kant lectures, Quine states that his “ontology is physicalist, rather than mentalist, from the start”. From an ontological point of view, this implies that we do not quantify over mental entities in our theory of world. Since physicalism implies that “[t]here is no change [...] without physical change” (1977, 281), we can get rid of mental entities by physicalizing them; every mental event can be identified with its corresponding bodily event. Or, as Quine expresses it in the Kant Lectures:

Dualism with or without interaction is redundant and reducible to a physicalistic monism, unless disembodied spirits are assumed. For, the dualist who rejects disembodied spirits is bound to agree that for every state of mind there is an exactly concurrent and readily specifiable state of the accompanying body. Readily specifiable certainly; the bodily state is specifiable simply as the state of accompanying a mind that is in that mental state. But then we can settle for the bodily states outright, by bypassing the mental states in terms of which I specified them. We can just reinterpret the mentalistic terms as denoting these correlated bodily states, and who is to know the difference? (IKL, 2)

Quine, in other words, does not need the mental in his theory of the world because mental states can simply be identified with bodily states.⁹

It should be noted, however, that although individual mental states are *ontologically* equated with individual bodily states, we can only *specify* these bodily states in mental terms: “[t]he bodily state corresponding to a mental one [is] only specified by reference to the mental state” (p. 3). Every time I think about Vienna, to use one of Quine’s own examples, there is a corresponding bodily state; but it is probably impossible to find a distinct physical mechanism M such that M is present if and only if I am thinking about Vienna. This implies that mentalistic descriptions (‘thinking about Vienna’) are *practically* indispensable for everyday life and for social science:

The mentalistic predicates, for all their vagueness [...] have long interacted with one another, engendering age-old strategies for predicting

and explaining human action. They complement natural science in their incommensurable way, and are indispensable both to the social sciences and to our everyday dealings. (1990, 72-3)

Furthermore, the idioms of propositional attitude (e.g. ' x believes that p ') are also practically indispensable for the process of language learning. In "States of Mind" (which is based on the first lecture of the Kant Lectures), Quine argues:

Take the observation sentence 'It's raining'. Tom is learning it from Martha by ostension. Martha's business is to encourage Tom in uttering the sentence, or in assenting to it, when she sees that he is noticing appropriate phenomena, and to discourage him otherwise. Thus Tom's mastery of the physicalistic sentence 'It's raining' hinges on Martha's mastery, virtual if not literal, of the mentalistic sentence 'Tom perceives that it is raining' [...] The handing down of language is thus implemented by a continuing command, tacit at least, of the idiom ' x perceives that p '. (1985, 325-6)

Quine's physicalism, in sum, is non-reductionistic. In the Kant Lectures, Quine cites Donald Davidson in describing his view as an "anomalous monism" and Daniel Dennett in claiming that his view combines "token physicalism" with "type dualism".¹⁰ Quine's physicalism is "not a reductionist doctrine of the sort sometimes imagined [...] the groupings of events in mentalistic terms need not stand in any systematic relation to biological groupings"; Quine's physicalism only entails that "there is no mental difference without a physical difference" (1977, 279).

5. The symptomatology of psychophysiology

Let me sum up what we have established thus far. I have argued that Quine is not a logical behaviorist. Quine does not claim that mental statements are acceptable if and only if they can be reduced to behavioral statements; he maintains that mental statements are acceptable if and only if they contribute to our overall theory of the world. In response to the question whether mental statements *do* contribute to our overall theory of the world, however, Quine's answer is negative. Although he admits that mental terms are *practically* indispensable because there are "irreducibly mental ways of grouping physical states and events" (1990, 72), his ontology is strictly physicalistic; because there is no change without physical change, we do not have to admit mental terms when we are "limning the true and ultimate structure of reality" (1960, 221).

Quine, in sum, rejects mentalism on an ontological level; our comprehensive theory of the world, if properly regimented, does not ontologically commit us to the existence of mental entities. But is this enough to qualify Quine as a behaviorist in psychology? In this final section, I will turn to the third type of behaviorism to answer this question: Skinner's epistemological behaviorism.

Thus far, Quine's account seems to be compatible with Skinner's. Skinner, like Quine, rejects logical behaviorism and combines a strictly physicalistic ontology with a distinction between private and public processes. Where Quine dismisses Hempel's positivist strictures on definition, Skinner rejects logical behaviorism by dismissing Bridgman's operationalist perspective (1945, 1984).¹¹ Most striking,

however, is the similarity between Skinner's and Quine's views about the ontogeny of mental terms. Where Quine argues that we learn to apply mental terms to private events "by extrapolation along similarity lines" (IKL, 14-5), Skinner argues that "verbal responses which are acquired with respect to public events may be transferred to private events on the basis of common properties" (1953, 259). Both Skinner and Quine, in other words, argue that our mental terms go beyond publicly observable behavior because we extrapolate along private channels.

But what about the most radical element of Skinner's behaviorism; what about the view that mental entities are explanatorily redundant? Again, there are good reasons to suppose that Quine's and Skinner's views are compatible. Skinner, we have seen, argues that we cannot genuinely explain behavior by referring to mental states: to say that Ahmed is eating an apple *because* he is hungry is merely to offer a redundant redescription: "A single set of facts is described by two statements: 'He eats' and 'He is hungry'" (Skinner 1953, 31). Quine seems sympathetic to Skinner's line of reasoning. For, in dismissing traditional theories of meaning, he uses almost the exact same argument: if we appeal to 'meanings' in explaining why we understand certain expressions, we are merely offering a "spurious explanation":

Meanings [...] purport to be entities of a special sort: the meaning of an expression is the idea expressed. [...] The evil of the idea idea is that its use, like the appeal in Molière to a *virtus dormitiva*, engenders an illusion of having explained something. (1951b, 48)

When we posit a *virtus dormitiva*, we do not explain the sleep-inducing quality of opium, when we say that someone is hungry, we do not explain his eating behavior, and when we posit a meaning, we do not explain why we understand an expression.¹²

Despite these similarities, however, it would be a mistake to classify Quine as an epistemological behaviorist. For although Quine and Skinner appeal to similar arguments in *dismissing* mentalism and logical behaviorism, they do not agree about the types of explanations that psychologists *should* seek. Quine, unlike Skinner, does not believe that behavior is ultimately explained by reinforcement histories. Like many of his colleagues in philosophy and psychology departments after the 1970s, Quine believes that behavior ultimately requires a *neural* explanation:

Behavi[o]rism, mine anyway, does not say that the mental states and events consist of observable behavi[o]r, nor that they are explained by behavi[o]r. They are manifested by behavi[o]r. Neurology is the place for the explanations, ultimately. But it is in terms of outward behavior that we specify what we want explained. (1978a, 10-1)

The importance of behaviorism is its insistence on shoring up mentalistic terms, where possible, by forging substantial links with observation. For a deep causal explanation of mental states and events, on the other hand, we must look not just to behavior but to neurology. (1980a, 26)

Behavior calls for explanation. And the explanation is going to be in the nervous system. It's going to be physiological. But the behaviorism goes in, in stating the problem that we are trying to solve. (1998, 94)

Quine, in other words, distinguishes between three domains in his philosophy of psychology: the mental, the behavioral, and the physiological.¹³ In talking about our

actions we often appeal to mental explanations; Paul cries because he has a toothache and Ahmed is eating an apple because he is hungry. The behavioral level specifies what needs to be explained: Paul's crying and Ahmed's eating. The physiological level, finally, offers the most fundamental explanation; the things, events, and processes referred to are ultimately states of nerves.

a mental state is a state of nerves, whether or not the neural mechanism is understood. Coming to understand that mechanism is what constitutes full explanation of the mental set, or event; and coming to understand the neural mechanism is likewise what constitutes full explanation of the behavior that manifests that mental state or event. (1980c, item 2999, my transcription)

In clarifying his view, Quine often compares mental states to diseases. Both mental states and diseases are recognized in terms of publicly observable symptoms; we know that Ahmed is hungry because he is eating and we know that Anne has jaundice because her skin looks yellow. Even if we know the physiological cause of the symptoms, we often keep specifying mental states and diseases in terms of its public and private symptoms. Ultimately, however, the symptoms require physiological explanations, even if in some cases we discover that there is no underlying physiological explanation at all:

Mental states are like diseases. A disease may be diagnosed in the light of its observable signs though the guilty germ be still unknown to science. Incidentally, diagnosis depends heavily on symptoms reported by the

patient; and such is the way, overwhelmingly, with the detection of mental states [...] In some cases a supposed disease has been ascribed, such as the vapors, that does not really qualify as a disease at all, there being no one germ or other uniform causal mechanism behind it. Now much the same is true of mental states. Often they are wrongly ascribed on inadequate evidence, such as false testimony, and sometimes a supposed mental state is ascribed that does not qualify for physicalism as a state at all. (IKL, 5-6)

The behavioral level, in sum, is the level at which mental states are *identified*—mental states that are ultimately equated with psychophysiological processes. As a result, Quine does not see “behaviorism as an *alternative to* psychophysiology” he sees it “as the *symptomatology* of psychophysiology” (1983, item 2851, my transcription and emphasis).¹⁴

Still, we should not conclude that there are *no* explanations at the behavioral level. Quine grants that Skinner is right that we might find regularities on the behavioral level as well.

Is this all we can say for behaviorism, and must all theory be left to the neurologist? Certainly not, as Skinner has long stressed. There is strong theoretical reason for shortcutting the neural wellsprings and seeking laws or uniformities within the behavioral level. The reason is that the uniformities and significant correlations in psychology have to do with the function or outcome of an animal’s movements, and are insensitive to differences in the anatomical details of the movements as long as they serve the same end. Running the maze is the thing, and no matter which

muscles are flexed in what order, or which neurons are fired to flex them. Attention to the anatomical detail of implementation would be counterproductive, obstructing the explanatory generalities and obscuring the woods with trees. (1989a, 348)

What we see here is another instance of Quine's anomalous monism. Because there are 'irreducibly mental ways of grouping physical states and events', we will not be able to find neurological explanations for all types of behavior. Often, behavior is classified functionally and we will require functional explanations if we want to justice to those classifications:

an explanation—not the deepest one, but one of a shallower kind—is possible at the purest behavioral level. One can hope to find, and I think one does find, behavioral regularities. In economics, for instance, you can formulate the concept of unemployment without reducing the phenomenon of economics to the behavior of individual people, which would be unproductive and chaotic. The instructive regularities occur at another level. In psychology that level is behaviorism. (1994, 94-5)

6. Conclusion

Let me, in conclusion, return to the question whether Quine can be classified as behaviorist in psychology. I have argued that Quine's position is surprisingly subtle. On the one hand, he believes that there are interesting explanations at the behavioral level, especially when we want to explain functionally clustered types of behavior. He

even goes as far as to state that *in psychology* instructive regularities occur at the behavioral level. On the other hand, however, Quine denies that behavior is ultimately explained in terms of reinforcement histories. Ultimately, mental states are *physical* states. Explanations at the behavioral level are merely ‘shallow’ explanations; we do not require such explanations when we are ‘limning the true and ultimate structure of reality’.

If we take into account these subtleties, it is perhaps not surprising that Quine almost always qualifies his views when he uses the term ‘behaviorism’. He speaks about “moderate behaviorism” (1980a, 26), about “behaviorism, in the form in which I find it acceptable” (1983, item 2851, my transcription) and about his “very moderate, and I would say [...] very reasonable behaviorism” (1998, 94). Furthermore, in private correspondence, Quine also admits that his behaviorism is “pretty moderate” (1980b, item 1004) and that he “perhaps [...] never qualified as a behaviorist” in the first place (1989b, item 177, my transcription).¹⁵

One thing Quine and the psychological behaviorists definitely have in common is their staunch rejection of mentalism. Still, even here Quine’s views are subtler than his critics make him out to be: Quine’s anti-mentalism does not imply that we do not require mental talk in everyday life and in the social sciences, that we cannot, to some extent, meaningfully talk about private events, or that all mental statements can be reduced to physical statements. Quine accepts that mental states are practically irreducible, that mental terms can be ‘extrapolated along private channels’, and that there are ‘irreducibly mental ways of grouping physical states and events’. Mental states, in sum, are a mixed bag; some of them are well understood physiologically, some of them can only be identified by their outward symptoms, and some, as we will probably discover in the near future, do not refer to a uniform physiological phenomenon at all.¹⁶

Literature

- Carnap, R. (1936). Testability and Meaning. *Philosophy of Science*, 3(4), 419–471.
- Carnap, R. (1956). The methodological character of theoretical concepts. In H. Feigl & M. Scriven (Eds.), *Minnesota Studies in the Philosophy of Science*. Volume 1 (pp. 38-76). Minneapolis: University of Minnesota Press.
- Føllesdal, D. (2011). Developments in Quine's behaviorism. *American Philosophical Quarterly*, 48(3), 273-82.
- Gibson, R. F. (2004). Quine's Behaviorism cum Empiricism. In Gibson (Ed.). (2004). *The Cambridge companion to Quine* (pp. 181-199). Cambridge, MA: Cambridge University Press.
- Graham, G. (2017). Behaviorism. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition). Retrieved from <https://plato.stanford.edu/archives/spr2017/entries/behaviorism>
- Guthrie, E. R. (1950). The status of systematic psychology. *American Psychologist*, 5(1), 97-101.
- Hempel, C. G. (1935/1949). *Analyse logique de la psychologie*. Translated by W. Sellars (1949). The logical analysis of psychology. In H. Feigl and W. Sellars (Eds.), *Readings in philosophical analysis* (pp. 373-384). New York: Appleton-Century-Crofts.
- Hempel, C. (1952). *Fundamentals of concept formation in the empirical sciences*. International encyclopedia of unified science, Volume II, number 7. University of Chicago: Chicago Press.
- Hempel, C. G. (1980). The logical analysis of psychology. In N. Block (Ed.), *Readings in philosophy of psychology* (pp. 14-23). Cambridge, MA: Harvard

- University Press.
- Hylton, P. (2007). *Quine*. New York: Routledge.
- Kemp, G. (2012). *Quine vs. Davidson: Truth, reference, and meaning*. Oxford: Oxford University Press.
- Putnam, H. (1957). Psychological concepts, explication, and ordinary language. *Journal of Philosophy*, 54(4), 94-100.
- Putnam, H. (1963). Brains and behavior. In R. J. Butler (Ed.), *Analytical Philosophy*. Second series (pp. 211-235). Oxford: Basil Blackwell.
- Quine, W. V. (1947). Letter to Morton White and Nelson Goodman, June 3, 1947. In White, M. (Ed.). (1999). *A philosopher's story* (pp. 338-340). University Park: Pennsylvania State University Press.
- Quine, W. V. (1951a). Two dogmas of empiricism. *Philosophical Review*, 60(1), 20-43. Reprinted in Quine (1953), pp. 20-46.
- Quine, W. V. (1951b). The problem of meaning in linguistics. In Quine (1953), pp. 47-64.
- Quine, W. V. (1953). *From a logical point of view*. Second and revised edition (1961). Cambridge, MA: Harvard University Press.
- Quine, W. V. (1954). The scope and language of science. In L. Leary (Ed.). (1955), *The unity of knowledge* (pp. 231-247). New York: Doubleday. Reprinted in Quine (1966), pp. 228- 245.
- Quine, W. V. (1959). Meaning. In Quine (2008a), pp. 163-165.
- Quine, W. V. (1960). *Word and object*. Cambridge, MA: The M.I.T. Press, Cambridge.
- Quine, W. V. (1966). *The ways of paradox and other essays*. New York: Random House. Revised edition (1976). Cambridge, MA: Harvard University Press.

- Quine, W. V. (1970). Philosophical progress in language theory. *Metaphilosophy*, 1(1), 2–19.
- Quine, W. V. (1975). Mind and Verbal Dispositions. In Quine (2008a), pp. 257-270.
- Quine, W. V. (1977). Facts of the matter. In R. W. Shahan & K. R. Merrill (Eds.), *American philosophy from Edwards to Quine* (pp. 176-196). Norman: University of Oklahoma Press. Reprinted in Quine (2008a), pp. 271-286.
- Quine, W. V. (1978a). The ideas of Quine. Interview by B. Magee. In B. Magee (Ed.), *Men of ideas* (pp. 168-179). London: BBC Publications. Reprinted in Quine (2008b), pp. 5–17.
- Quine, W. V. (1978b). Otherworldly. *New York Review of Books*, November 23, 1978, 25(18):25. Reprinted as ‘Goodman’s *Ways of worldmaking*’ in Quine (1981), pp. 96-99.
- Quine, W. V. (1980a). Sellars on behaviorism, language, and meaning. *Pacific Philosophical Quarterly*, 61(1-2), 26-30.
- Quine, W. V. (1980b). Letter to Laurence D. Smith. December 8, 1980. W. V. Quine Papers (MS Am 2587): Box 036, Item 1004. Houghton Library, Harvard University.
- Quine, W. V. (1980c). Opening remarks for question session with psychologists. Autograph manuscript. March 14, 1980. W. V. Quine Papers (MS Am 2587): Box 103, Item 2999. Houghton Library, Harvard University.
- Quine, W. V. (1981). *Theories and things*. Cambridge, MA: Harvard University Press.
- Quine, W. V. (1983). Calcutta lectures: The mentalistic heritage. January 1983. W. V. Quine Papers (MS Am 2587): Box 100, Item 2851. Houghton Library, Harvard University.
- Quine, W. V. (1985). States of mind. *Journal of Philosophy*, 82(1), 5-8. Reprinted in

- Quine (2008a), pp. 323-326.
- Quine, W. V. (1987). Indeterminacy of translation again. *Journal of Philosophy*, 84(1), 5-10. Reprinted in Quine (2008a), pp. 341-346.
- Quine, W. V. (1989a). Mind, Brain, and Behavior. In A. J. Brownstein (Ed.). *Progress in behavioral studies* (pp. 1-6). Mahwah: Lawrence Erlbaum Associates. Reprinted in Quine (2008a), pp. 347-351.
- Quine, W. V. (1989b). Letter to Howard G. Callaway. September 28, 1989. W. V. Quine Papers (MS Am 2587): Box 006, Item 177. Houghton Library, Harvard University.
- Quine, W. V. (1990). *Pursuit of truth*. Revised edition (1992). Cambridge, MA: Harvard University Press.
- Quine, W. V. (1994b). Interview with Willard Van Orman Quine. Interview by L. Bergström & D. Føllesdal. *Theoria*, 60(3), 193- 206. Reprinted in Quine (2008b), pp. 69-81.
- Quine, W. V. (1998). There is always a further step. Interview by O. Brill. In Quine (2008b), pp. 82-97.
- Quine, W. V. (1999). Response to Segal. A. Orenstein & P. Kotatko. (Eds.), *Knowledge, language and logic* (pp. 417-418). Dordrecht: Kluwer Academic Publishers.
- Quine, W. V. (2008a). *Confessions of a confirmed extensionalist and other essays*. Edited by Føllesdal, D. & D. B. Quine. Cambridge, MA: Harvard University Press.
- Quine, W. V. (2008b). *Quine in dialogue*. Edited by Føllesdal, D. & D. B. Quine. Cambridge, MA: Harvard University Press.
- Roth, P. A. (2006). Why there is nothing rather than something: Quine on behaviorism, meaning, and indeterminacy. D. Jacquette (Ed.), *Philosophy*,

- psychology, and psychologism* (pp. 263-287). Dordrecht: Kluwer Academic Publishers.
- Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review*, 52(5), 270-277.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Skinner, B. F. (1969). *Contingencies of reinforcement*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1984). Author's response. *The Behavioral and Brain Sciences*, 7(4), 572-581.
- Verhaegh, S. (2014). Quine's argument from despair. *The British Journal for the History of Philosophy*, 22(1), 150–173.
- Verhaegh, S. (forthcoming). *Working from within: The nature and development of Quine's naturalism*. New York: Oxford University Press.
- Verhaegh, S. (ms.). The behaviorisms of Skinner and Quine: Genesis, development, and mutual influence.
- Watson (1913). Psychology as the behaviorist views it. *Psychological Review*, 20(2), 158-177.
- Zuriff, G. E. (1985). *Behaviorism: A conceptual reconstruction*. New York: Columbia University Press.

¹ See Quine (1959, 163): “If ideas did exist we’d have to disregard them”; (1970, 4): “A language is mastered through social emulation and social feedback, and these controls ignore any idiosyncrasy in an individual’s imagery or associations that is not discovered in his behavior”; and (1999, 417): “my linguistic behaviorism [...] disciplines data”. For an overview of Quine’s linguistic behaviorism, see Gibson (2004), Roth (2006), and Føllesdal (2011).

² Methodological behaviorists have offered a wide range of objections against the study of mental states, processes, and images: they have argued that introspective data is subjective, that introspective experiments cannot be replicated, that the connection between private events and verbal behavior is unreliable, and that theories based on introspective data cannot be falsified. See, for example, Guthrie (1950, 99): “what appeals to me as the outstanding aim and the requirement of science [is] its public character—its foundation in human communication, not merely in the private experience of individuals. Scientific observations must be repeatable by others—there can be no science until there are men using a common language”.

³ John B. Watson, for example, defends a combination of methodological and ontological behaviorism. In “Psychology as the Behaviorist Views it”, for instance, Watson argues both that introspection is unreliable and that “thought processes are really motor habits in the larynx” (1913, 177). B. F. Skinner’s “radical behaviorism”, on the other hand, combines an epistemological argument against mentalism with a functional analysis of behavior. See Zuriff (1985) and Graham (2015) for taxonomies of behaviorism that are more closely aligned with positions that have been actually defended by twentieth-century psychologists.

⁴ Of course, Carl Hempel (and many other logical positivists) would later abandon the translationist view summarized here. See section 3.

⁵ Carnap (1936, §8) characterizes partial definitions as follows. Let Q_3 be a mental predicate, let Q_1 and Q_4 describe experimental conditions which have to obtain in order to find out whether or not Q_3 applies, and let Q_2 and Q_5 describe possible results of the experiments. Then Q_3 can be introduced as a new predicate in one's language by statements like R_1 and R_2 :

$$(R_1) \quad Q_1 \rightarrow (Q_2 \rightarrow Q_3)$$

$$(R_2) \quad Q_4 \rightarrow (Q_5 \rightarrow \neg Q_3)$$

Definitions of this form are *partial* definitions because Q_3 is only specified relative to a set of experimental conditions Q_1 and Q_4 . See Verhaegh (2014; forthcoming, section 2.2).

⁶ It is probably for this reason that both Carnap and Hempel would later also abandon partial definitions. See Hempel (1952, 32) and Carnap (1956, 68). Again, see Verhaegh (2014; forthcoming, section 2.2). Indeed, in a third reprint of “The Logical Analysis of Psychology”, Hempel admits that even his weakened version of logical behaviorism is too strong: “Since then, I have come to think that this conception requires still further broadening, and that the introduction and application of psychological terms and hypotheses is logically and methodologically analogous to the introduction and application of the terms and hypotheses of a physical theory” (1980, 14).

⁷ See also Quine's *Calcutta Lectures* (1983), written a few years after the *Kant Lectures*: "Full operational definition of terms is not [...] to be demanded, even for rigorous science; but if a term is seriously deficient on that score, then it needs to justify itself by contributing theoretically to the systematic explanation of observed events. Such was the contribution of the notion of molecules in explaining the behavior of gases by the mechanics of moving particles" (item 2851, my transcription).

⁸ In the *Kant Lectures*, Quine makes the same point without explicitly referring to mental terms: "Empiricism or positivism at its radical extreme would aspire to a completely operational lexicon. The reasonable line rather is one that plays two values one against the other. There is a premium on perceptual criteria: the fuller the better, other things being equal. But there is also a premium on structural simplicity and the other related qualities, whatever they are, that make for a satisfactorily explanatory scientific theory. A term that promises well in this latter way can be excused its remoteness from perceptual criteria. A judicious weighing of these two values is what is called for" (IKL, 19).

⁹ This argument goes back to beginning of Quine's philosophical career. Indeed, in the early 1950s, Quine already claimed that a "physicalist ontology has a place also for states of mind" because any "inspiration or a hallucination can [...] be identified with its host for the duration" (1954, 230).

¹⁰ Although Quine adopts Davidson's label 'anomalous monism', the latter's version of the thesis differs from Quine's in some crucial respects. See Kemp (2012).

¹¹ For a comparison between verificationism, operationalism, and partial definitions, see Zuriff (1985, ch. 3). In the *Immanuel Kant Lectures*, Quine also characterizes logical behaviorism as a type of operationalism: "this extreme of behaviorism would

be excessively restrictive. It would be as unreasonable as an unswerving insistence in scientific theory on what Bridgman called operational definition. Indeed it would be a case of that” (p. 5).

¹² See also Quine (1947, 339-40): “Frege, Carnap, Lewis, and the rest seem to derive from those shadowy **entities [attributes, propositions, and meanings] the same smug illusion of clarity that Toletus did from his substantial forms, and Moli[è]re’s physician from the *virtus dormitiva***”.

¹³ Quine first appeals to some such distinction in “Mind and Verbal Dispositions”, although he (somewhat misleadingly) speaks about “three levels [...] of *reduction*” there (p. 253-254, my emphases). I thank Robert Sinclair for bringing this passage to my attention.

¹⁴ After the Immanuel Kant Lectures, Quine often reused the disease analogy. See, for example, Quine (1980c, item 2999; 1983, item 2851; 1985; 1994; 1998). It is surprising, however, that Quine never credited Putnam for the metaphor. After all, Putnam often used the same analogy in arguing for his functionalist philosophy of mind. See, for example, Putnam (1957; 1963).

¹⁵ These qualifications start to appear from the late 1970s onwards. This suggests that Quine changed his mind on this issue, perhaps even in preparing the Immanuel Kant Lectures. For a discussion of the development of Quine’s views on behaviorism, see Verhaegh (ms.).

¹⁶ An early draft of this paper was presented at the 2018 APA Central Division symposium on Quine’s Immanuel Kant Lectures. I thank Robert Sinclair, Gary Ebbs, and the audience in Chicago for their valuable comments and suggestions. This research is funded by The Netherlands Organisation for Scientific Research (NWO, Grant 275-20-064). My archival research at the W. V. Quine Papers was funded by a

Kristeller-Popkin Travel Fellowship from the *Journal of the History of Philosophy*, by a Rodney G. Dennis Fellowship in the Study of Manuscripts from Houghton Library, Harvard University, and a travel grant from the Evert Willem Beth Foundation.