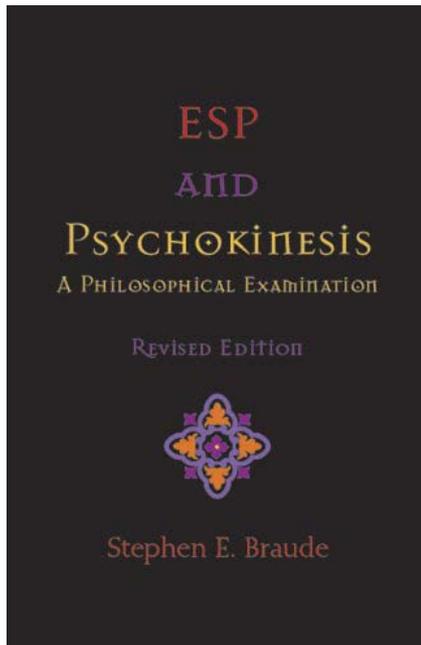


BOOK EXCERPT



Stephen E. Braude

ESP and Psychokinesis
A Philosophical Examination
(Revised Edition) ↗

Paperback, 246 pages
Brown Walker Press ↗
Copyright © 1979, 2002

ISBN-10: 1581124074
ISBN-13: 9781581124071

pp. 145–176
Republished by permission

Psi and the Philosophy of Mind

1. *Anomalous Monism*

I want now to consider how the objections raised above against the *ET* theory bear on some traditional issues in the philosophy mind — in particular, how they bear on psychophysical identity theories.^A It is common practice these days to distinguish two general forms of the Identity Theory. One asserts lawlike correlations between *types* of mental states and *types* of physical states. The *ET* theory is a version of the type-type form of the theory; and although parapsychologists, psychologists, and other scientists dabbling in the study of consciousness often cling somewhat naively to type-type theories, more sophisticated philosophers typically abandon this form of the Identity Theory in favor of *token-token* versions of the theory.

^A The energy-transfer theory of telepathy (*ET*, for short) is discussed in Section 4d of Chapter B, pp. 123–140. In its attempts to reduce telepathic processes to a certain kind of mechanical physical process between agent and percipient, it postulates the existence of specific physiological states corresponding to semantic properties of thoughts yet fails to specify physical regularities corresponding to semantic regularities between thoughts. — Ed.

According to that form of the Identity Theory, every mental state-token is a physical (let us say, brain) state-token, but there need be no lawlike correlations between brain state-types and mental state-types. For example, my thought that S is Φ will be identical with a brain state of mine, but there need be no lawlike correlations between thoughts of the kind that S is Φ and any sort of brain state at all. Davidson (Davidson, 1970) calls this position *anomalous monism* (hereafter, *AM*). His view is monistic, since he takes all mental events to be brain events. But the relationship between the domains of the mental and the physical remains anomalous — that is, there need be no correlating laws for the two domains.

Thus, some might argue that although my arguments against the *ET* theory weigh against type-type versions of the Identity Theory, they do not touch *AM*. But that contention, I want now to argue, is wrong. More specifically, I will first explain why some of my arguments against the *ET* theory apply directly against *AM*. Then I will consider further arguments against *AM*, including the argument that *AM*, rather than escaping the pitfalls of type-type Identity Theories, actually *presupposes* type-type correlations between the mental and physical. Therefore, I will argue that *AM*, like the *ET* theory, is disguised nonsense.

One of the principal reasons the *ET* theory collapses is that (as I argued earlier and will argue from a somewhat different point of view in the next section) mental states can only be characterized *positionally* — that is, with respect to a certain way of fitting events into a larger sequence of events. For example, if something specific is going on in A 's mind,¹ (say, the occurrence of an image of a collie), we don't yet know what A 's mental state is (that is, what his image is an image *of* for him) unless we know how A 's mental image is incorporated into a sequence of relevant events. The same image can represent different things in different contexts.

Now *AM* considers A 's mental state to be some particular brain state. But if we take A 's physiological state at a given time, and the collie image allegedly correlated with it, and if we place them in different contexts, we might properly characterize A as thinking about dogs generally or about Lassie in particular. For example, in one context A might have his image (or be in the corresponding brain state) in response to the instruction, "Think about dogs;" in the other, the instruction might have been, "Think about Lassie." But if it is possible for the brain state corresponding to A 's mental image to represent different things in different contexts, then apart from any context the brain state itself represents no more than the mere image itself considered apart from

¹ I use this example for the sake of simplicity. But we should remember that a person can be in a mental state m without there being any particular *experience* corresponding to m . For example, every time I brush my teeth, I can be described properly as remembering how to brush my teeth, even though no occurrent mental states of mine (i.e., no memory experience) could properly be described as my remembering how to brush my teeth. Similarly, while discussing football with a friend, I might reach into my trouser pocket for my house key; and although there need be no occurrent inner episode of mine that could be characterized as a memory (or belief) experience concerning my house key, I could properly be described as remembering where my key was (or believing that my key was in my pocket).

any context. So to the extent that *AM* construes mental states to be nothing but mere states of persons — in other words, to the extent that *AM* rests on a static and topological, rather than a dynamic and functional, characterization of mental states — it has reached the same dead end as the *ET theory*. *A*'s mental states cannot be characterized simply in terms of what is happening solely within *A*. But then contrary to *AM*, a particular mental state in *A* cannot be a particular physiological state of *A*.

A similar objection weighs against the extended static and topological view of thinking that identifies a particular physiological state *plus* a certain specific set of contextual conditions. As we have seen, from a topological point of view, that context is as functionally ambiguous as the physiological state conjoined with it. If we placed it in a different and *wider* context, it might be correlated with an entirely different mental state. Moreover, as I point out in detail in the next two sections, there are in fact no privileged or context-independent preferred descriptions of contexts. Hence, there are no privileged or context-independent preferred ways of integrating a particular event into its surrounding context.

This objection to a static or non-functional characterization of mental states may be linked with an objection advanced by Wittgenstein, 1955a, 1955b, and thereafter by Malcolm, 1977 — namely, that we apply radically different concepts of *duration* to brain states and to mental states. Neural firings and biochemical changes in the brain are events whose duration, occurrence, or temporal boundaries can be explicitly and fairly precisely measured. In general, however, beliefs, desires, intentions, and so forth — the sorts of things whose instances (according to *AM*) are particular physiological states (or physiological states plus physical contextual states) — cannot be measured in this way. That is not to say that mental states do not last or have duration. Rather, it is to say that many (if not most) mental states — both occurrent states and dispositions — are simply not the sorts of events or processes for which the making of such measurements is conceptually appropriate in a clear way (see Wittgenstein, 1955b: paras. 45, 81ff, and Malcolm, 1977: pp. 257-260). But this shows that mental states generally and brain states are quite different sorts of things, and it robs of all antecedent plausibility the suggestion that instances of one are instances of the other.

Proponents of *AM* actually concede that there may be no way of mapping the *predicates* applicable to mental states onto predicates of *any* language used to describe the brain states with which they are allegedly identical. For example, the predicate “__ is a belief that Jones is a coward” or the predicate “__ is a thought about Lassie” may map onto no predicate in the language of even the most advanced brain science. Therefore, *kind* terms appropriate to mental states may fail completely to correlate with *kind* terms appropriate to brain states. But as Bruce Goldberg has argued, proponents of *AM* tend not to notice an important corollary of that fact (see Goldberg, 1977). Let Ψ be the portion of language used to describe mental states, and let Φ be that portion of language used to describe physical states — in particular, states of the brain. The anomalous monist concedes that the predicate expressions of Ψ need not map onto the predicate expressions of Φ . But as Goldberg points out, there can then be no reason to suppose that the *referring* expressions of Ψ map onto those of Φ .

Suppose we correctly say of Smith that he is in a certain mental state-token m and that m is an instance of believing that Jones is a coward. The anomalous monist grants that there may never be an expression in the language of the physical sciences coextensive with “_ is a belief that Jones is a coward.” In the view of the physical sciences, that psychological kind term might not mark off a genuine kind; hence, no predicate of Φ needs to correspond to this predicate of Ψ . But if so, then we have no reason to suppose that we will ever form an expression in the language of the physical sciences having the same *referent* as “ m .” If the *kinds* of phenomena describable in Ψ fail to map onto the kinds acknowledged in the physical sciences, then we have no reason to suppose that the *objects* to which the predicate expressions of Ψ apply are (or map onto) the objects to which the predicate expressions of Φ apply. If physiology and psychology distinguish different kinds of phenomena — if they do not divide the world into kinds or types in anything like the same way — they may also have entirely different criteria of individuation within their respective domains. And in that case, their domains may not divide up into members, or tokens, in anything like the same way.

But then the anomalous monist can claim no grounds for identifying tokens in one domain with tokens in the other. Having conceded that kinds of mental states might not correspond to kinds of physical states, the anomalous monist must grant that parsings of physical states need not correlate *at all* with parsings of mental states. So, in principle, no member of the domain of mental states needs to correspond (or be identical) with any member of the domain of physical states. Thus, *AM*'s assertion of token-token identity between these two domains becomes entirely arbitrary — or else merely a form of wishful thinking. (If, after this, there is any remaining appeal in the view that mental state-tokens and physical state-tokens are identical, I suspect it is due to the unrecognized fact that *AM* presupposes type-type correlations between the mental and the physical. Thus, anomalous monists tend not to realize that they actually do *not* concede that mental state-types might not correlate with physical state-types. I discuss this issue below.)

Hard-core anomalous monists might be unmoved by the foregoing considerations. They might argue that those objections show only that we may never get very far with actual psychophysical theories. That is, we might never be able to achieve an actual reduction of the language Ψ to the language Φ . Even so, they might say, *AM* advances a substantial metaphysical claim that the objections so far offered leave untouched. That claim runs more or less as follows.

(α) S is in mental state-token $m \Rightarrow$ there is some brain state-token b of S such that without b , S would not be in m .

(α) is much weaker than the claim of token-token identity. In fact, (α) need not even be a *monistic* claim. Obviously, it does not assert an identity. Rather, it asserts a kind of causal dependence of m on b , and thus it seems to be an epiphenomenalist claim. Moreover, b is a necessary but not a sufficient condition for the occurrence of m . That conforms with points raised earlier, since (as we have seen) what mental state a person is in is at least partly a function of context. Because physical or physiological structures

are functionally ambiguous, considered only with respect to their topological features, no brain state as such has any intrinsic single function. Thus, (α) represents a substantial shift away from the already weak identity claim we initially attributed to the anomalous monist.

But we can see why (α) would appeal to some philosophers. It still anchors mental states in physical states, and although it does not identify the two, it still denies mental states a life of their own. In other words, (α) seems to be a commendably modest physicalist position, replacing identity between the mental and physical with mere causal dependence of the former on the latter.

But let us look more closely at (α) . In my opinion, it has very little to commend it. According to (α) , when S is in mental state-token m , S 's being in m is explained in terms of S 's being in some brain state-token b necessary for the occurrence of m . But now we must ask proponents of (α) an embarrassing question: In explaining m , does it matter what *type* of state, physiologically speaking, the token b is a token of? If the answer is yes — as I will argue must be if (α) has anything of interest to say — then (α) is really a type-type view in disguise, and my earlier arguments apply against it. But if the answer is no — if it does not matter what kind of brain state the token b is — then b can be *any* old state. All (α) asserts in that case is that there must be *some* brain activity for a person to have mental states — perhaps simply that we must be *alive* to have thoughts.

Now it is true that my earlier objections do not touch that very weak position. Although so construed, (α) is not a trivial claim — for it seems incompatible with certain dualistic and especially spiritistic views according to which mental processes are autonomous — it is no longer an *explanatory* thesis. It cannot explain in physiological terms why m occurred rather than some other mental state. Nor would it help explain why m is just the kind of mental state it is. To do that in physiological terms, we must revert to a type-type correlation view, which we've seen is unsatisfactory. In any case, if it does not matter what the physiological features of brain state-token b are, then (α) promises no physiological understanding of why m occurred (rather than a token of a different type), or of what, physiologically speaking, it means to have a thought of type M . If Smith is thinking that Jones is a coward, partisans of (α) can say no more than that Smith has that thought because he is in *some* brain state or other. They cannot tell us why, physiologically, Smith has that particular thought rather than another thought, or what it is, physiologically speaking, to have (say) a thought that Jones is a coward (or even what it is to have any thought about cowards).

Nor can (α) provide the foundation for a theory having any predictive utility. In fact, if b can be any brain state at all, then (α) provides no physiological basis for the science of psychology. We would have no way of describing or predicting a person's mental state in physiological terms. Understood in this exceedingly weak way, (α) is actually contrary to the original spirit of materialism, which was initially conceived as a program for explaining why a person was in one mental state rather than another, or what it is (in physiological terms) to be in certain types of mental states, or how (on the basis of one's physiological states) we might predict what a person's mental states would be. But these explanatory and predictive features of a materialist theory may be

secured only if there are type-type correlations between the mental and the physical.

Actually, these last remarks help to illuminate a fatal presupposition of the stronger token-token identity theories. The problem is that there is no justification whatever for positing an identity between mental state- and brain state-tokens unless type-type identities are presupposed. For example, consider a case where *A* is properly described as thinking about collies. According to token-token identity theories, that thought-token is identical with some brain state-token *b*, identified (let us say) with respect to some set *T* of topological features. But at this point we must ask: Can numerically different brain state-tokens having just this set *T* of topological features correlate with different mental states? Can some brain state-token *b'* which, like *b*, is individuated with respect to its being an instance of brain states having the set *T* of physiological properties, occur with some thought *other* than a thought about collies? If the token-token theorist answers yes (as may be expected of someone who denies type-type correlations) – if, that is, there can be a many-one relationship between mental state-types and the brain state-type individuated by the set *T* of topological properties – then there is *no reason* to identify *A*'s thought-token about collies with *b*, since state *b'*, topologically identical with *b*, can be correlated with a different thought. Thus, the identification of *A*'s thought about collies with *b* seems completely arbitrary.

It appears, then, that the identification of mental state-tokens requires that we presuppose type-type correlations between mental states and brain states. Only thus may we avoid the self-defeating admission of the possibility of a one-many relation between brain states of the physiological type *T* and types of mental states. But then, since it presupposes a type-type theory, *AM* proves, after all, to be subject to just the criticisms raised earlier against the *ET* theory of telepathy.

2. *The Myth of the Internal Mechanism*

I suspect that many readers will feel uneasy about the stand I have taken against the various forms of the identity theory. Perhaps they rebel against the *a prioristic* nature of my arguments. Perhaps they find it odd that purportedly empirical theories should be rejected on non-empirical grounds. If so, I must remind these readers that scientific theories rest – probably inevitably – on philosophical presuppositions; in that sense, they are really philosophical theories in disguise. (We will see this also in II.B, in connection with the so-called theory of synchronicity, although to call that view a *scientific theory*, or even a theory, may be unreasonably generous.) Thus, the various forms of the *ET* theory rest on many philosophical assumptions, despite their superficially empirical or non-philosophical character. I have been at pains to criticize some of these – for example, the thesis that structures of the brain display their function, or that all thoughts of kind Φ share some common property or set of properties (a Platonic essence) in virtue of which those thoughts are Φ , or that there is a set of necessary and sufficient conditions for something to be a Φ -thought. A theory that rests on such assumptions rests on nothing, and no empirical arguments are needed to demonstrate its inadequacy.

Another probable source of dissatisfaction is harder to dispel. Indeed, I have to work hard to dispel it myself. The problem arises by way of a very natural (but mistaken)

assumption — so natural, in fact, that it undergirds the entire field of cognitive psychology, generative linguistics, almost all areas of the brain sciences (in fact, a good part of the biological sciences generally), and a great deal of philosophy. But before trying to state the assumption, let me offer some illustrations of the contexts in which we tend to appeal to it. I want the reader to appreciate from the start how fundamental and sweeping a revision our view of human beings and human behavior must undergo, once we abandon the assumption.

Imagine that we are trying to teach someone how to conjugate Spanish *ar*-verbs. Imagine, further, that our student has been unable to do more than memorize the conjugation of the verbs already shown her; she is unable to conjugate any new *ar*-verb presented to her. But after some further instruction, she finds that, for the first time, she can indeed conjugate new *ar*-verbs and has thus acquired an understanding she did not possess previously. Or, imagine that we are trying to teach a child the color red by teaching her to distinguish red things from non-red things. And suppose that in this process the child suddenly and clearly understands what to do. The assumption we tend to make in such cases is that something has happened, *within* the person, between the time she did not understand and the time she did. In other words, we tend to explain the newly-acquired understanding in terms of some associated structural modification of the person. Most would say it was a change in her brain.

Or imagine that Jones, who for many years had been a callous and frivolous person, suddenly has a religious experience that completely transforms his character, so that he becomes a loving, serious, and God-fearing person. We tend to assume that the change in Jones's character must be due to some corresponding physiological change *within* Jones himself (specifically, a change in his brain) produced by the religious experience. We assume that something relevant about the way Jones is put together is different from what it was before. (I say "relevant" because, after all, Jones must be changing physiologically all the time, whether or not he undergoes a religious conversion.)

Or imagine that one day I see a person whom I recognize as my old friend A, whom I have not seen in 10 years. We ordinarily assume that I could not remember A were it not for something *within* me (a memory trace) caused by my previous acquaintance with A, and that I could not recognize A after 10 years were it not for some specific physiological process occurring within me (the process of recognition).

These assumptions, I believe, are entirely false and seriously misleading. Of course I cannot hope to demonstrate in a detailed way how very nearly every aspect of even just cognitive psychology (to take one general field) is based on a mistake, by examining why these assumptions or their variants are unjustified in memory theory, recognition theory, learning theory, and so on. The general mistake appears in different guises in these and other areas. I will try instead to analyze the error in its most general and abstract form, and then show how it applies in certain cases. I then leave it to the reader to extend the arguments to other relevant areas in the study of human beings.

The assumption I want to attack may be called the *Principle of the Internal Mechanism*

(hereafter, *PIM*). In stating this principle I want, for simplicity, to insist on a certain terminological restriction. Hereafter, when I refer to the class of *mental states* or *events*, I want to exclude “raw feels” generally and sensations or images considered merely phenomenally. Although there is good reason to think that the points I will be making about thoughts, beliefs, memories, dispositions, instincts, and so on can be made, *mutatis mutandis*, about such things as pains and itches (see Lurie, 1973), it would needlessly complicate matters to discuss this entire range of experiences. I will therefore confine myself to more robust or intentional mental states. I realize that this does not mark off a crisply circumscribed range of phenomena. But I don’t believe a neat characterization is even possible.

With this in mind, the principle I wish to criticize may be formulated as follows.

(*PIM*) It is possible to explain (through some empirical theory) why *S* is in (occurrent or dispositional) mental state *m* by reference to some corresponding physiological structure or mechanism *b* identical with, or causally responsible for, *m*.

The *PIM* is the backbone of those areas of science whose entire program rests on the supposition that, in principle, we can discover mechanisms for at least some human cognitive abilities, or thoughts, beliefs, and so on. Should the *PIM* be false, numerous academic disciplines would thus turn out to have no foundation. Notice, moreover, that the *PIM* is not merely a version of a physicalistic identity thesis. Nor is it an epiphenomenalist principle. Rather, it is presupposed by identity theorists, epiphenomenalists, and even some dualists (e.g., those who insist on the autonomy of the mental in at least some cases, but who permit some physicalistic/mechanistic explanations of mental states).

Although I will argue that the *PIM* is false, that principle is by no means crazy, and in fact it is quite seductive. The reason for this can be seen clearly in the application of the *PIM* to memory. We tend to assume that I could not possibly remember my friend *A* after 10 years were it not for something (a trace) in me — let us say a modification of my brain — caused in the past through my association with *A*. Otherwise (we suppose), in remembering *A* after all these years we would have causation over a temporal gap. To explain how I remember *A* after 10 years, we must (it seems) say more than that I used to know *A*. When I remember *A*, therefore, it is presumably because something *persisting* in me links my present memory of *A* to my having known *A* 10 years earlier.

But notice that the appeal to a memory trace is not designed solely to explain how I happen to be in a specific mental state — namely, the remembering of *A*. It is intended to explain how such remembering is *possible* in the first place. The purpose of positing the existence of a trace (or of the structural modification allegedly associated with understanding a new concept, acquiring a new disposition, etc.) is to explain how a certain kind of mental state could *ever* occur — as though without such structural modifications, the mental events in question would be impossible. Without a trace in me — without a physiological representation of *A* — no experience that I now have *could* be a memory of *A*. For example, if after 10 years I call up a mental image of *A*, that image could not be a memory of *A* if it lacked the appropriate causal history linking my present experience to my previous acquaintance with *A*. My image would not count as

a memory unless there was some trace in me, caused by my previous acquaintance with A, activating that image. Mental images might be possible without such activation, but they cannot then be instances of remembering.

But what if traces *cannot* explain how we happen to be in that particular mental state whose very possibility that structural modification allegedly insures? In other words, what if the positing of traces cannot possibly explain any particular mental state-token of remembering? If in the explanation of any particular instance of remembering the role to be performed by the trace turns out to be an impossible role, a role nothing could fulfill, then we will have to reconsider our supposition that we must posit such a modification just to demonstrate that states of remembering were possible in the first place.

And that is what I will try to show. I will argue that any theory in which mental states get explained by reference to specific associated physiological structures or mechanisms that produce them is a theory that must rest on one or another false presupposition — specifically, either the Platonic assumption that mental states of type Φ share some common property (or set of properties) in virtue of which they are of type Φ , or else the unintelligible assumption that a physiological structure can be intrinsically functionally unambiguous. My discussion will focus almost exclusively on the case of memory. But the arguments apply, *mutatis mutandis*, to other cognitive functions.

The errors underlying the notion of a memory trace² (or more generally, the notion of the structure or mechanism associated with any occurrent or dispositional mental state) are basically just the ones discussed earlier in connection with the *ET* theory and anomalous monism, although the fine details of the mistakes there and here differ somewhat. We have already observed that a given brain structure, identified physiologically with respect to a set T of topological properties, is functionally ambiguous. That is, we have seen that topologically identical brain states can have different functional or representational properties, depending on how we position them with respect to the background conditions or context in which they are embedded. We also saw that the pretended explanatory power of the *ET* theory and *AM* requires the nonsensical assumption that some kind of structure (purely physiological, or physiological *cum* external) is functionally unambiguous — that is, that there is *not* a one-many relationship between the type T of the physiological state and the types of mental states to which it may be correlated. And in analyzing the *ET* theory, I remarked that in order to identify physiologically some mental state m as being of type Φ , we must suppose that there are necessary and sufficient conditions for a mental state to be

² Recently, some sustained attacks on trace theory have appeared in print, those offered by Bursen, 1978, Heil, 1978, and Malcolm, 1977. They explore in much greater detail (and, especially in Malcolm's case, from a somewhat different point of view) the sorts of criticisms I offer here. If these works were to have the effect they deserve on the intellectual community, many neurophysiologists and neuropsychologists might find themselves unemployed. Bursen is particularly blunt on this point, charging (correctly) that trace theories are scientific theories only on the surface and that, at bottom, they require that memory is magical.

of that type. In effect, there must be some specifiable property (or set of properties) that all mental states of type Φ share, and in virtue of which a given state is of that type. But that requirement proved that *ET* theory rested on an elementary form of Platonism.

Now, the situation is no different if we suppose that physiological structural modifications (of the brain) underlie the acquisition of new memories, concepts, or dispositions. For instance, we can easily expose the Platonic presuppositions in the notion of a memory trace. Recall that a thought of a certain kind (say, the thought of an old rabbi) may take an indefinite number of different forms, having nothing more of relevance in common than the fact that they are of that kind. When I am thinking of an old rabbi, I may be remembering a verbal description, or having an image of a person, or perhaps a more complicated image (say, of a rabbi conducting a bar mitzvah), or perhaps I am just presenting to myself the words “old rabbi,” or perhaps I am imagining those words sung by the Mt. Sinai Hospital Chorale. And of course each of these categories will have innumerable variations as well. For example, the verbal description or the images can differ subtly or dramatically from one another in their details.

But the same is true of any instance of remembering my friend *A*. There is no one form for that mental state, no one general feature or set of features that any token of remembering *A* must have. For example, remembering *A* might consist in nothing more than having any of a number of mental images or different feelings; or it might consist of certain bits of behavior, such as verbally describing *A*, or telling anecdotes about *A*, or reacting in any number of different ways upon meeting *A* after 10 years. For example, there might be a shock of recognition. But there might be no shock or feeling of recognition, or no experienced lapse of 10 years at all. On seeing *A* my sole reaction might simply be to say (or think), “You owe me \$50.” Moreover, an instance of remembering *A* might consist of nothing more than thinking (or talking) about things associated with *A* — say, cupcakes, or the Washington Monument. But if so, then it is clear that one cannot specify conditions for something’s being a remembering of *A*.

Rememberings of *A*, then, can take an endless number of different forms and need have nothing of importance in common except being instances of remembering *A*. But behind the diversity of possible forms of remembering *A* is supposed to be some modification in me (or my brain), something that was not present in those days before I first met *A*. That is the trace of *A*. Its existence is supposed to help us understand any particular instance of remembering *A*. We are to explain the occurrence of any such instance by linking it causally to the activation of the appropriate trace.³

³ How the “right” trace gets activated is a problem that memory theorists tend not to see clearly. They tend to suppose that it can be explained by reference to structural isomorphism between the inner state (the trace) and some other external thing (like a picture of *A* or *A* himself). But, as we saw in the case of the *ET* theory, that idea rests on the untenable claim that isomorphism between two things is intrinsic to those things, and that rests on the unacceptable idea that some structures are functionally unambiguous and that their topologies determine their own representational properties or rules of projection. (How fantastic that idea is becomes clear if we recall that a brain state does not represent or

But if instances of remembering *A* need not have any relevant properties in common save that of being rememberings of *A*, the trace serves no explanatory function. We can always ask: How can the activation of some particular state *b* cause the variety of things counting as rememberings of *A*? Where do we find a causal regularity? How can the same state causally explain both my having a mental image of *A* and my simply saying to *A*, “You owe me \$50?” If rememberings of *A* need have no relevant properties in common, then why suppose there is some single trace to whose activation all those rememberings are causally linked? And if there is no one thing (no common property or set of properties) to which we may link the trace, the trace serves no causal role after all. There is no causal regularity to appeal to, because there is no specifiable regularity in the range of possible effects of the trace. But if there is no such regularity, if *anything* experienced or done by a person can, in the right context, count as a remembering of *A*, then the posited *A*-trace explains none of those things.

Trace theorists might respond by saying that there are different traces for different *kinds* of rememberings of *A*. But that won't help; in fact, it starts trace theorists on an endless regress. Those subsets of rememberings are no more linked by a relevant common property (or set of properties) than were the various things that originally counted simply as rememberings of *A*. For example, mental images of *A* no more need such a set of common properties than do the different possible ways of imagining an old rabbi. The image could be of *A* as a young man or old man; it could be an image of a portrait of *A*, of *A*'s front, or profile, or *A*'s eyes, or face; the images could be rough images (like stick drawings) or more detailed (like photographs); they could be in color, black and white, or perhaps just in shades of blue; and so on. And so the question again arises: If the members of some subset of rememberings of *A* need have no relevant properties in common (save that of belonging to that subset), why suppose they are causally linked to the same thing?

Here a certain objection is likely to occur to some readers. They might argue that paradigm cases of causal connections exhibit the same diversity of effect that memory traces have, and these paradigms (they would say) are not thereby rendered suspect. For example, the triggering of dynamite can have various kinds of effects. We can use dynamite to blow up a bridge, a house, or a person; and each of those classes of effects will cover a great variety of phenomena. Dynamite explosions may be loud or muffled, messy or neat, etc. And here, the diversity of possible effects of dynamite is no reason for saying that the triggering of dynamite is not the cause of such effects.

Although this is an important objection, it misses a certain crucial point about the putative effects of memory traces. Despite their diversity, the various things caused by the triggering of dynamite are still explosions. But manifestations of the activation of memory traces needn't have *anything* of interest or importance in common except being instances of a certain type of memory. (They will, of course, have in common an infinite number of irrelevant or trivial properties, such as, being things, or being self-

resemble a person *A* in the way that even a photograph of *A* represents or resembles *A*.) In the next chapter and in the following section, we will examine yet another aspect of this mistake — namely, the unacceptable idea that things or events have a preferred parsing.

identical.) In advance of any actual triggering of dynamite, we can say that in all probability it will produce an explosion. But we cannot say in advance *how* a person will display a certain kind of memory once the appropriate trace is activated. For example, depending on the sorts of things associated with A, thinking about a Schubert piano sonata, or tuberculosis, or loose jello, might count as manifestations of remembering A. So the alleged parallel does not obtain.

At this point trace theorists are faced with a fatal dilemma. It arises because they must somehow avoid the unacceptable Platonic assumption that all instances of a mental state-type Φ share a common property (or set of properties) in virtue of which they are of type Φ . Grasping one horn of the dilemma, they could concede that there may be a one-many relationship between the trace and the kinds of things that count as rememberings of A. But in that case, since the rememberings of A are not themselves connected by a set of necessary and sufficient conditions for being a remembering of A, the trace (as we have seen) can have no explanatory power. If there is no specifiable property (or set of properties) that something must have in order to be a remembering of A (and not something else), then there is nothing in any actual remembering of A to link the trace to.

Seizing the other horn of the dilemma, trace theorists can avoid positing a one-many relationship between the trace and the possible forms of its causal consequences by saying that each remembering of A (i.e., each token of remembering A) is causally linked to its own unique trace. In that way, they can avoid supposing that there is a property (or set of properties) common to rememberings of the same type. But that alternative still robs the trace of any explanatory power. If each remembering of A has its own unique associated trace, then first of all the positing of traces does not explain the general ability to remember A, because rememberings of A may take different forms. But one reason theorists initially posited the existence of traces was to explain how we could ever have such a general ability. Thus, positing a different trace for each manifestation of an ability fails to explain how the ability itself is possible.

Moreover, the second horn of the dilemma assigns to traces the functional uniqueness we found to be unintelligible in the context of the ET theory. Memory traces are now linked one-one with particular manifestations of remembering A; the trace is the causal antecedent of only one specific token *m* of remembering A — it *cannot possibly*⁴ be causally linked to any different human experience or behavior (including other rememberings of A having no relevant properties in common with *m* and even non-rememberings of A). Trace theorists must say that *m*'s trace cannot possibly be causally linked to mental state-tokens of different types (or even different tokens of the same type), because otherwise they would allow for the one-many relationship between the physical and the mental which the retreat to this horn of the dilemma was supposed to avoid. If the trace can cause other sorts of human responses or behaviors (including

⁴ The type of possibility here is empirical rather than logical. That is, the trace theorist is committed to saying that as a matter of empirical fact (i.e., given the laws of this world) the trace cannot have more than one associated effect, although there may be possible worlds in which some trace *t* has an effect it doesn't have in this world.

non-memory responses), then it does not explain any of the different effects it may have. But if the trace is to be causally linked with only one token of one mental or behavioral state-type (as trace theorists must now say), then we are back to the untenable position that a (brain) structure or mechanism can have one and only one functional role, no matter what the context is in which it is embedded.

As a matter of fact, there are numerous places where trace theorists are committed to the existence of intrinsically unambiguous structures or mechanisms (no matter what the details of their theory). I have already suggested that the appeal to unambiguous structures is inevitable at the point where trace theorists must explain how the appropriate or right trace gets activated in any particular case of remembering. But the same difficulty appears much earlier in the theory, simply in virtue of supposing that some state of the brain does one and only one thing – say, uniquely represent *A*. After all, if my *A*-trace could represent someone or something besides *A*, then it may cause me to remember something besides *A*. And in that case, we could not explain why I remember *A* and not someone or something else by appealing to the trace.⁵ But (as I observed earlier) a brain structure (of whatever kind) can no more represent in virtue of features intrinsic to it than can a photograph (imagine *A*'s photo appearing in a book under the heading “Human Being,” or “Caucasian,” or “Man with Receding Hairline”).

The trace's representational properties are supposed to be exclusively a function of its topological properties, those properties of its physiological or physical structure in virtue of which the trace represents what it does and not some other thing. But what a thing *a* represents is never strictly determined either by its own properties, or by its properties conjoined with those of some context *c* in which it is embedded. Context *c* will be just as intrinsically ambiguous as representing object *a*.

Consider: If *a* and *c* are both physiological structures (like brain states), both will be functionally ambiguous for the reasons already discussed. And if *c* is something like a societal or conversational context, it will be ambiguous in a different way. How *c* should be characterized will always be relative to some point of view from which the characterization of *c* makes a difference. It has no intrinsic characterization. We must somehow describe the context, or pick it out, if only to say what it is that *a* must be conjoined with in order to represent what it does in *c*. But how we describe the slice of life we are taking to be a context is not something forced on us by features of that bit of life itself. We must make a decision about how to parse and label it, and there will always be different ways of doing that, each one reflecting some distinctive point of view (recall the familiar philosophical example of describing the same episode as

⁵ Trace theorists might protest that the *A*-trace is causally linked with remembering *A* in just those cases where the brain is in some *further* state *s*, and in this way they might try to show how the admittedly functionally ambiguous trace could represent *A* and not something else. But that move obviously pushes the same problem back a step, since the *A*-trace and *s* taken *together* are as functionally ambiguous as the trace alone. If trace theorists deny this, they are again positing the existence of intrinsically unambiguous structures. But if they agree, then the new state (*A*-trace & *s*) no more explains remembering *A* than did the original trace.

turning on a light, or flipping a light switch, or alerting a burglar). But since nothing about *c* demands that it be described in one and only one way (that is, since there is no inherently privileged or context-independent preferred description of *c*), we do not, by conjoining *a* with *c*, avoid the intrinsic ambiguity of *a* itself. In fact, even if *c* were another brain state, it would have no intrinsic or context-independent preferred parsing. (I will pursue that point in the next section.)

In order for one thing to represent another, then, it must occupy some position in a bit of history; it cannot represent anything independently of a context. But what gets represented is never rigidly fixed by the bit of history within which we consider it. Because there is no one context-independent preferred description of any context in which *a* is positioned, there will also not be one context-independent preferred description of *a*'s relations to its context. But if there are in principle different ways of relating *a* to its surrounding context, then although *a* needs a context in order to represent anything, what we take it to represent, or whether it represents anything at all, will depend on which relations we choose. Independently of any context or reason for choosing one set of relations over another, there is no preferred choice. However, when we adopt the *PIM*, we commit ourselves to the position that there do exist intrinsically unambiguous physiological representations — for example, traces of objects (as of *A*) or representations of concepts (like that of the color red). But (to stick with memory traces) if even a photo of *A* does not intrinsically represent *A*, then no brain structure can or will.

At this point, some might try to salvage the *PIM* by shifting from memory traces to instincts. Surely, they might say, the *PIM* is not false for all cases, since even relatively simple organisms (like insects) have instincts, and instincts may plausibly be identified with (or causally traced back to) certain associated (and presumably persisting) structures or mechanisms in the insect's nervous system. If that is plausible in the case of insects (they might ask), why not in the case of humans?

I would say, however, that this is *not* plausible even in the case of insects. The problems here do not concern the representational properties of internal structures; but, as with memory traces, the claim rests on a form of Platonism, because it requires that instinct-mechanisms be connected causally with a class of effects united by some non-trivial, relevant *common property*. Take, for instance, the instinct to survive. That instinct seems to be exhibited by creatures on all levels of the evolutionary scale. But how is it expressed? Among insects, it is expressed in the search for food, in the construction of nests, in avoidance behavior under certain conditions, and even in aggressive behavior under others. In other words, even in its most rudimentary manifestations, the instinct to survive can take a variety of forms, which (like remembering *A*) need not be linked by any relevant common property (or set of properties) except the property of being an instance of the instinct to survive. There need be no set of necessary and sufficient conditions for something's being an instance of the instinct to survive. But then the structures or mechanisms whose activation allegedly causes the instinctive behavior (like the memory trace whose activation allegedly explains remembering *A* or the ability to remember *A*) do not explain the

instinct after all. In neither case is there some regularity or property common to the effects of the structure or mechanism to which we can causally link that structure or mechanism.

Furthermore, just as memory-trace theory faced a fatal dilemma once it was forced to concede that effects of the trace needn't share any relevant common properties, a similar fate confronts instinct-mechanisms. Suppose we identify some putative instinct-mechanism *i* with respect to a set *T* of topological features. We may now ask: Is the relationship between *i* and the kinds of causal consequences it may have one-many or one-one? If it is one-many, then since the kinds of consequences needn't share some Platonic essence or common property for which we can state necessary and sufficient conditions, there is no causal regularity we can specify and then *i* has no explanatory power after all. But if we say the relationship is one-one, then since instincts are manifested in different kinds of ways, we must concede that *i* no longer explains the instinct as a whole. Moreover, to causally link *i* to only one specific token of the instinct is to take the untenable position that *i* is functionally unambiguous — that is, that *i* cannot correspond to or causally explain more than one token of one type of behavior. But since what kind of behavior a bit of behavior is will be a function of the way we embed it in a set of background conditions, this position works no better for *i* than it did for the memory trace. In neither case can we classify the effect of the structure simply in terms of the structure's properties, and independently of the properties of the situation in which that effect is produced.

Applying these conclusions to the range of physiological mechanisms posited in cognitive psychology (e.g., in learning theory, recognition theory, or in theories analyzing beliefs and thoughts), some important morals emerge for the philosophy of mind. Of these, two are of immediate benefit. First, appeals to a persisting physiological structure or mechanism — to explain *abilities* or *dispositions* to react or behave in certain ways (ascribed to persons or other organisms) — will not work, because nothing can be specified among its possible effects to which the structure or mechanism may be causally linked. And second, appeals to a physiological structure or mechanism — to explain, not a class of experiences or behaviors, but only tokens of some type of experience or behavior — will fail as well. Either the structure or mechanism may be identified with (or causally linked to) tokens of various types of states or different kinds of tokens of one type — in which case nothing is explained (since those diverse states need have no relevant properties in common); or it can be identified with (or causally linked to) one and only one token of only one type of state — in which case the putative structure or mechanism is an impossible object.

From the point of view of any viable explanatory theory, abilities, instincts, dispositions, thoughts and beliefs are all in the same boat. They all have manifestations which (like my ability to remember *A*) need not be linked by any relevant specifiable common property (or set of properties). But then, since we have seen that no structure or mechanism explains an ability, disposition, etc., whose manifestations are linked by no relevant common property (or set of properties), the only causal antecedents to which we need appeal in explaining abilities, dispositions, etc. — if we need appeal to any at

all — will be episodes in the organism's history. Once we grant that changes in a person's cognitive abilities need not correspond to any internal structural change in the person, we can admit that these abilities are simply brute facts about the organism for which certain kinds of "why?" questions are no longer appropriate. If so, then to explain how I can remember *A*, we need appeal only to certain events in my past. Similarly, when we consider changes in disposition (e.g., Jones's conversion from being frivolous to being serious), we need say no more than that the change makes sense relative to a certain episode in Jones's life (his religious experience, say). Here, then, we may abandon the search for a further explanation in terms of something happening within Jones himself (e.g., within his brain). Nothing of that sort could explain the change.

Many readers will no doubt recoil at my suggestion that we stop seeking explanations of abilities, dispositions, etc., in terms of underlying physiological structures or mechanisms. For example, some may protest that, since we obviously do not remember *every* event in our past, we must do more than explain present remembering in terms of past experiences. In particular (they would say), we must appeal to a persisting physiological modification produced by the relevant past event. Now granted, more *can* be said than that certain events in my past enable me to remember *A* now. But in light of the foregoing objections, I would argue that we should not appeal to internal mechanisms. What we *can* say (for instance) is that certain events in my past, involving *A*, were particularly intense or important at the time and that I tend to remember intense or important events. (In fact, even on memory-trace theory we must explain why some past events and not others produce traces.) Or perhaps I am simply a person able to do certain things rather than others. I might remember *A*'s phone number but not his name, because I have an ability to remember numbers and not names. Here, we explain my remembering *A*'s phone number in terms of other regularities of mine, like my ability to remember numbers generally. But that general ability is a fact about me for which no structural explanation (say, in terms of memory traces) exists or can exist.

Another objection, bound to be raised sooner or later, is that studies in cerebral lesions and the like show that mechanisms for at least some abilities and dispositions do exist. After all (my opponent would say), we can alter or impair cognitive abilities and even one's character by altering certain parts of one's brain. These (they would contend) are cases in which changes in a person's mental states are explainable as due to corresponding physiological changes.

But it seems to me that such cases really show very little. The loss of a cognitive ability after a brain lesion does not show that there is (or was) some brain mechanism for the lost ability. All it shows is that the persistence of a person's abilities, dispositions, etc. requires maintaining the general integrity of the organism. The falsity of the *PIM* is compatible with the fact that we can alter a person's occurrent or dispositional mental states by changing the person physically. All the rejection of the *PIM* demands is that we stop our search for specific underlying structures or mechanisms by which to explain the abilities, dispositions, etc., in question.

To see this, suppose that after damage to a certain very specific region of the cortex, a person *S* loses his ability to remember numbers, though his other mnemonic abilities (like his ability to remember names and faces) remain unimpaired. Would this show that a certain brain structure or mechanism *b* in *S* had been the cause of *S*'s ability to remember numbers? I would say not, and to see why we need only review the points raised against the trace theory.

Suppose, for argument's sake, that we identify a brain structure or mechanism *b* whose activation is allegedly responsible for *S*'s ability to remember numbers. And suppose that we specify *b* in terms of a set *T* of topological properties (i.e., that we give a "hardware" description for *b*). Consider now what sorts of things may be manifestations of remembering numbers. *S* might demonstrate his ability to remember numbers by speaking, writing, or by pointing to numbers (say, on a chart), or by pointing to a collection of things the sum of whose members is the number remembered, or (say) by tapping his foot in order to imitate a horse counting with his hoof. So first of all there is an indefinite and presumably unlimited number of ways *S* could demonstrate his ability to remember numbers. Moreover, a given token of that ability — for example, pointing to a number on a chart — need not serve as a token of remembering a number. What in one context counts as an instance of remembering a number may in another context count as an instance of something else. For example, *S* might point to the only red number on the chart after being asked to identify which number is red. In that case, our familiar dilemma appears again. The brain structure or mechanism *b* can, when activated, cause an indefinite number of different states and behaviors, which are not intrinsically tokens of remembering numbers, and which need not be linked by any relevant common property (even when they are instances of remembering numbers) save that of being instances of remembering numbers.

Thus as before, no particular token of *S*'s remembering a number belongs to a specifiable causal regularity in the range of *b*'s possible effects. So the association of *b* with the token is arbitrary. And as we have seen, the only way to avoid linking *b* with tokens of different types of abilities, or with different kinds of tokens of the same type, is to make *b* the sort of thing which can cause one and only one token of one type of effect. But since any of *b*'s effects may be instances of any number of types of effect, this amounts to assigning an impossible role to *b*.

Therefore, that we can alter a person's mental states by altering the organism does not demonstrate the possibility of physiologically mechanistic explanations of a person's mental states. Whatever structure or mechanism we isolate can never correspond in the right way to the mental state or class of mental states it allegedly explains. To repeat: If the structure or mechanism can be identified with (or cause) tokens of various types of mental states or different tokens of one type, then since those various states need have no relevant properties in common, the structure or mechanism (considered simply as a bit of hardware) has no explanatory utility. And if it can be identified with or cause only one token of only one type of mental state, then we are attributing a role to the structure or mechanism which nothing can fulfill.

By the way, I am not saying that we must *supplement* appeals to episodes in a person's

life with appeals to brain mechanisms. Rather, I am saying that appeals to brain mechanisms *add nothing* to our reference to the episodes in the person's history. After all, we have seen that appeals to brain structures or mechanism confer on those putative items either no explanatory role, or else an impossible role.

Nevertheless, rejection of the *PIM* does not force us to conclude that a person's physiological constitution is *irrelevant* to what that person's mental states are. For example, we needn't deny that brain size and complexity of brain topography (e.g., convolutions) are proportional to intelligence (whether or not that is a fact). Rejection of the *PIM* requires only that we stop looking for something in the organism, some bit of hardware, which (or the activation of which) accounts for the particular (occurrent or dispositional) mental state we want to explain. We must abandon the search for a certain kind of explanatory *unit* (like a brain structure) more fundamental than the organism itself.

Organisms do indeed differ from one another, and some of their physiological differences might be causally connected with some of the psychological differences between them. For example (to take a simple case), we can admit that the difference between Einstein's intelligence and that of my cat is the result of the difference in their brains. But to admit *that* is not to grant that there are physiological mechanisms that explain the various features of, say, Einstein's intelligence. Moreover, as far as changes in a person's character or mental states are concerned, things can happen to an organism in light of which the alteration of its behavior or mental life makes sense. And these may be highly specific things, like Jones's religious experience, my previous acquaintance with A, or damage to a certain region of the brain. But to suppose that we *must* look further *within* the organism for specific explanatory structures or mechanisms is just to commit the very mistakes I have been laboring to expose.

The supposition that we *must* look further is no more than an article of faith, based (it seems to me) on a reluctance to admit that certain sorts of macroscopic facts might be ultimate facts, not capable of explanation by further dissection and analysis. There is no compelling reason to suppose, either in this area of inquiry or others, that explanation by analysis should be able to continue indefinitely. Now without trying to provide a recipe for discovering where, in any inquiry, we should stop the process of explanation by analysis, I submit that a perfectly good place to stop is just where further analysis can only rest on false or absurd presuppositions.

When we begin to take this possibility seriously, we will be drawn to a somewhat different view of life than the one that makes the *PIM* seem plausible. And that inclination will be strengthened when we also consider that nature has no intrinsic structure which it presents to us, but that we instead parse slices of history in different ways for different purposes (see II.A.3 and II.B for discussions of this). I don't believe I am up to the task of describing this alternative view of life in great detail (since I am still in the process of discovering what it is). But I might be able to indicate how it forces us to revise our conception of what thoughts or mental states must be.

To be seduced by the *PIM*, we must take a certain broad view of what thoughts or mental states are, and of what it means to endow such things with content or meaning.

In general, we must believe that thoughts and mental states are certain kinds of *things*; that (in principle at least) we could formulate an ontology in which thoughts or mental states had a place. There are many ways of doing this. For example, we could say that thoughts are brain events, or immaterial mental structures, or even propositions (construed as abstract logical structures). And of course these views can be held with varying degrees of sophistication. We might take thoughts to be brain events, while also granting that our mentalistic vocabulary is theoretically defective — that mentalistic terms do not divide the class of mental states in any philosophically or scientifically acceptable way (see, e.g., Davidson, 1970; Dennett, 1978). From that viewpoint, it is our working set of mentalistic categories that fails to provide us with well-defined or definable terms; the domain of the mental may itself remain capable of being divided into well-defined categories. Moreover, since thoughts or mental states are individuated with respect to their content or meaning (e.g., we speak of the thought or belief *of*, or *that* ___), it is easy to treat thoughts and mental states as concrete things. We need only suppose that, at least in principle, we can specify the content or meaning of those things clearly and exhaustively. Although (we might say) we still lack sufficiently refined theoretical machinery for the task, in principle nothing prevents us from getting it done.

I want now to suggest that this general outlook is profoundly and entirely mistaken. We have already considered one reason for this judgment — namely, that mental states do not fall into classes united by any relevant common property (or set of properties). Classes of mental states are simply not neat in a manner amenable to scientific treatment. But unlike those who take this lack of neatness as a characteristic only of our mentalistic *vocabulary* (or conceptual categories) — hence, a problem that can be solved by appropriate conceptual development — I have been suggesting that this lack of neatness is a fundamental feature of mental *phenomena*, and that our categories are already somewhat neater than the phenomena to which they apply.

When, in the inescapable process of abstraction, we arrive at such categories as remembering (believing, wishing, etc.) that *p*, we inevitably foster the illusion that the range of phenomena under discussion is (in principle) isolable and specifiable — if not by our current stock of categories, then by more adequate future categories. And that illusion is reinforced by our actual success in using our categories to describe or predict human behavior. We tend to forget that the process of abstraction that generates our supply of mentalistic concepts is merely a prerequisite for communication. But there is no reason to suppose *a priori* that any mentalistic vocabulary is anything more than a communicative tool. In particular, there is no reason to suppose *a priori* that either our mentalistic vocabulary or the domain to which it applies must be anything more than fuzzy.

In fact, when we recall that inner episodes, bits of behavior, and sentence-tokens have content, meaning, or significance only with respect to the way we position them in a certain parsing of life, and when we recall further that the position of an experience, bit of behavior, etc., with respect to its context is not inherently a clear-cut matter, it becomes more difficult still to cling to the view that the domain of the mental —

mental states as well as their contents — is anything but fuzzy. The range of possible parsings or descriptions of a context (we must remember) is as unlimited as the range of possible perspectives we can take on life itself. There is simply no one correct way (or one context-independent correct way) of positioning an experience or bit of behavior against its context.

Moreover, the language in which we make our parsings and specify context or meaning is subject to the same intrinsic ambiguity as the thing whose meaning or content we are specifying. In individuating mental states (or behaviors, sentences, etc.) there can never be a point at which we fall back on something intrinsically unambiguous. (In my view, this difficulty has not been sufficiently appreciated by those who think we can give the meaning of a sentence *S* in terms of another sentence — for example, one stating the truth conditions for *S*.)

Although this is not the whole story, once these points begin to sink in, the meaning or content of a person's state or behavior will no more seem to be a discrete or in-principle-specifiable thing than (say) the *humor* or *sensitivity* of a bit of behavior. For example, a person's behavior is seen as humorous when we integrate it in a certain way with its surrounding history. A remark is funny in relation to something that preceded it, though not to any and everything that preceded it. A remark's humor is a *positional* property of the remark,⁶ not something inherent in it. It is entirely possible that two people *A* and *B* regard remark *r* as funny — but not in exactly the same way — even though we could properly describe them both as having understood the joke. Given *A*'s history, interests, and his own brand of perceptivity, *r* might have nuances, overtones, or a kind of complexity and richness that it does not have for *B*, or even for the joke-teller *C*, given the different perspectives they bring to the situation. Although these different sets of nuances, etc., are not intrinsic to *r*, they are also not mere appendages to the (essential) humor of *r*; indeed, the humor of *r* is nothing without them.

In fact, there *is* no essential humor to *r*, even in just those contexts in which *r* is funny. First of all, *r* is funny only in virtue of how we position it in a context; positioned in another way it might not be funny at all. Moreover, since there is an endless number of equally legitimate or correct perspectives, attitudes, etc. that we can bring to our understanding of *r*, there are endlessly different ways of construing *r*'s content. Thus, the question “what is the humor of *r*?” has no single — much less a uniquely correct — answer. Since *r* has no humor independently of some perspective on its context, and since no perspective is intrinsically preferable to any other, not only is *r*'s humor not the sort of thing that can be exhaustively specified, it is not even *one* thing. Certainly, it is not simply a function of any intrinsic features of *r*.

The same, I would say, is true of the meaning or content of an inner episode or bit of

⁶ I owe this terminology to Bruce Goldberg, and I prefer “positional” to “relational,” because the former term suggests that the humor of a remark depends on how we position it in its context and on what we bring to our understanding of the context. By contrast, the latter term suggests more that a remark's humor is a static and intrinsic feature of it and its context. Moreover, the term “relational” has a long philosophical history, and I wish to avoid as far as possible whatever misleading associations its use may entail.

behavior (as I have been urging ever since I.B.4.d). The meaning or content of what goes on in a person's mind or of what a person does is, like the humor of *r*, a positional property of that thing. What the inner or outer episode's content is depends on the context in which it occurs. But since life does not come pre-parsed for us — that is, since there are many different legitimate ways to characterize any slice of history — any bit of human behavior or human experience may be meaningful in different ways relative to different characterizations of (or perspectives on) its context. So, because experiences and bits of behavior have no content or meaning independently of context and perspective on context, and because no perspective from which behavior and experience are meaningful is intrinsically preferable to any other, this content or meaning is not something our behavior and experience simply *have*. Like the humor of *r*, content or meaning is not just one thing, and it is not the sort of thing that can be exhaustively specified.

If we take this seriously, we must construe human history, not as a series of events with a manifest structure or content, but rather as an intrinsically undifferentiated flow of experiencing and behaving that can be parsed and related to one another in various ways, and also reveal different of their aspects from the points of view associated with these different parsings and positionings. Just as, from the perspective of one such parsing, we can individuate a certain episode and see it as humorous, we can individuate the things people do or experience and view them as significant relative to other things individuated within the same perspective. But since the content or meaning of an inner or outer episode is a positional property and not something it has intrinsically, we must realize that mental states in general are not context-independent elements in the manifest structure of history; there is no such manifest structure. Mental states exist only relative to some parsing and characterization of the flow of life.

There are, of course, inner mental episodes; I am not denying their existence! And of course some of these can be dated and measured. But such episodes belong to an ongoing flow of inner activity that itself has no preferred parsing. Moreover, a person's mental life consists of more than inner episodes, and when we take that seriously, it becomes easier still to abandon the view that mental states and their contents generally are episodes or kinds of discrete things. For example, my belief (or memory) that I have a foot may be expressed in the putting on of my shoes or in my walking from one place to another (even when my inner episodes concern other things, like last night's football game). There needn't be any isolated state which we can identify as the belief (or memory). Similarly, my remembering that I have an appointment at noon may be manifest in my rushed behavior all morning, and not necessarily in a recurring memory state. My fear of being psychologically abused may be expressed in my hostile behavior, and not in a specific fear state.

Within the ongoing stream of life we identify various regularities and relate them meaningfully to one another or to specific episodes. Of course, that identification presupposes some way of individuating the elements within the stream. But that will only mark a selection from the endless possible ways of parsing and describing history.

Moreover, these regularities themselves have histories, and some will persist longer than others. For example, dispositions may change — as in Jones’s conversion from a frivolous to a serious person. And sometimes we can identify what makes the new regularity intelligible — for example, Jones’s religious experience. In other words, Jones’s religious experience may be meaningfully related to an identifiable pattern in Jones’s behavior. That is not to say that the religious experience *activates* diverse but related behaviors in Jones, as (say) the triggering of a memory trace is supposed to activate diverse things counting as memories of a certain type. It is only to say that we can *explain* (in one sense of this term, roughly equivalent to “render intelligible”) identifiable behaviors or behavioral patterns in someone’s life, not by locating activating mechanisms or structures associated with them, but simply by linking these things with others so as to form meaningful patterns relative to our present interests. This, I am inclined now to think, may be as far as we can ever hope to push causal explanations of human behavior.

I realize these remarks have been sketchy, possibly provocative, and worthy of further exploration. But I cannot pursue these deeper issues here. I will add only that the idea that all of life and history is an intrinsically undifferentiated flow with an infinite number of aspects (no one of which is absolutely preferable to any other), is rather similar to what mystics have been saying for a long time. But to regard this view as mystical is to miss an important point. The alternative to this position is to say that nature has an intrinsic structure — some built-in division into elements, and that there are fundamental relationships between these elements that their arrangement somehow displays unambiguously. In short, the alternative to this position is really a kind of Tractarian logical atomism. And that, I submit, is a far more fantastic viewpoint than the one I offer in its place.

3. *Some Comments on Recent Theoretical Trends*

Some of the objections raised earlier against the *ET* theory, anomalous monism, and the *PIM* generally apply also to several theories attracting attention in parapsychology. Although many parapsychologists regard these theories as representing promising lines of conceptual development, in my view they are nothing more than complicated variants of crude atomistic psychological theories dressed up in contemporary scientific jargon. None represents even a slight conceptual advance over the most primitive atomistic, static, or non-functionalist theories of consciousness, whose errors have been addressed in the two previous sections of this chapter and in I.B.4.d.

One theory which some parapsychologists regard as promising is the attempt by Karl Pribram to analyze memories, perception, and other sorts of conscious states in terms of *holographic* states of the brain (Pribram, 1971, 1977, 1974). Many who are interested in parapsychology regard Pribram’s work as a revolutionary advance in the study of consciousness, one that might finally lead the way to the long-awaited development of a satisfactory scientific theory of the paranormal. But this seems to me to be completely unjustified. For one thing, Pribram’s memory theory is simply a *memory-trace* theory with a twist. Pribram is apparently oblivious to the difficulties inherent in the notion of a memory trace. He seems not to realize that traces have no explanatory value (as I

have shown in the last section), and he seems also not to understand why it is unacceptable to characterize mental states solely in terms of their topological properties or as the non-positional states of persons (an error to which my discussion of the *ET* theory, anomalous monism and the *PIM* have been devoted).

Somewhat innocently, Pribram thinks that we can clear up certain long-standing puzzles about memory and human cognitive functioning by suitably complicating the nature of the memory trace (and similar physiological counterparts), ascribing to it (and these counterparts) the properties of a hologram. For example, he thinks that his holographic model helps to explain how memories may be retained even after excision of part of the cortex. In his view, information contained in the brain (like information in a hologram) is contained in any of its parts. But since the content of a mental state — the information Pribram takes to be contained in the brain — is not solely a function of brain physiology, his theory cannot be more than an elaboration of an idea already found to be unacceptable. It does not matter how complicated we make the brain states supposedly identical with (or the causes of) specific mental states. No structures of any kind can do the job that Pribram requires of memory traces. No structure determines its own (or a unique) function, which is why thoughts cannot be characterized topologically. Hence, Pribram fails to avoid the *ET* theory's fatal error of analyzing function in terms of structure, and (more specifically) that of the *PIM*. Moreover, the question of *where* in the brain memories are located (before or after brain surgery) arises only if we assume that memories must be somewhere in the brain, or simply somewhere. But once we give up the idea that memories (or mental states generally) must be specific, concrete *things*, we will no longer feel compelled to assign them a location. The basis for Pribram's entire neuropsychological program thus comes to nothing.

Of course, parapsychology is a field that invites ambitious and often reckless speculation, and Pribram's holographic model has inspired various cosmic extensions. The line most commonly taken is that the universe as a whole can be viewed holographically — that is, that every part of the universe contains the information of every other part (see, e.g., Anderson, 1977). In this way, some hope to avoid the problem of explaining, by means of a quasi-perceptual mechanism or mechanism of transmission, the awareness of remote information. According to this neo-Leibnizian picture, information about remote parts of the universe is already present in every part. So instead of worrying about the *transmission* of information from remote places (e.g., thoughts, in telepathy), this approach posits something like *resonances* between similar structures — just as Pribram and his followers hope to explain the activation of appropriate memory traces (as in associative memory) by means of resonances between structurally isomorphic states of the brain. But this view, like the *ET* theory, is hopelessly and fatally tied to the *PIM* — and in particular, to the view that the content or meaning of a brain state (i.e., its representational properties) is a function of its structure. But brain structures do not determine their own function (or a unique function); and whatever representational properties a brain state has will depend upon (though they will not be strictly determined by) how we position the state in a surrounding context. Thus, resonances between structures cannot explain, for instance, how my thoughts give me

access to the similar thoughts of remote individuals, or how a thought of one kind makes me have an associated thought of a similar kind (cf. my discussion of isomorphism, in I.B.4.d).

Much in the same spirit, E.H. Walker has attempted to analyze the content of conscious states information-theoretically, as constructs of *bits* — as though bits were fundamental or atomic constituents of mental or brain events rather than conventionally defined components, and as though the content of a mental or brain event is simply a function of its intrinsic structure (see Walker, 1974, 1975). Many in parapsychology regard Walker's efforts as important and pioneering. But it seems to me that his view is nothing more than an old-fashioned sense-datum theory stated in the language of the electrical engineer and thus, like Pribram's, it is basically unintelligible.

One reason for this is that, like Pribram, Walker supposes that there can be structures in the brain (or mind) that have or determine one and only one function. Of course, that is the error explored in detail in I.B.4.d and II.A.1. The other main reason goes hand in hand with this one and involves another mistake, which I have so far only hinted at. One who assumes that brain states have a structure that uniquely determines their content posits the existence of (impossibly) functionally unambiguous structures. However, he must also suppose that when a brain state is (or causes) a thought, the brain state's structure can be determined independently of determining what that thought is. More generally, views like those of Pribram and Walker presuppose that nature has a structure intrinsic to it (whether macroscopic, microscopic, subatomic, or logical) just waiting to be discerned.

To see why that is a fatal presupposition, consider the following. Suppose that Jones is describing his new home to Smith, and suppose that he tells Smith that the tree by his house is taller than the house itself. Let us keep things as simple as we can, and let us try to characterize Smith's subsequent thought that Jones's tree is taller than his house. Let us suppose that what happens in Smith when he has his thought is that he has an image in his mind of a tree rising above a house. Of course thoughts need not be images. However, we tend rather easily to think that elements of images are clearly manifest. (For those, like Walker, who think that *perceptual* experiences have a manifest structure, we need alter this case only slightly, by construing the image of the tree and house as the mental image Smith has when he *looks* at Jones's house.)

But what *are* the elements of the image? We are assuming that the thought Smith has when he has the image is the thought that the tree is taller than the house. So we might say that the elements of the thought (image) are a tree and a house, as relata of the relation “_ is taller than _.” But, as we have observed on numerous occasions, this image, set in a different context (that is, positioned differently in a sequence of events), could have functioned quite differently from the way it does here, so that we could have characterized Smith as having a different thought. Jones might not have told Smith that the tree was taller than the house. He might have said other things instead, which caused Smith to have the same mental image of the tree and house. But in that case we might reasonably have ascribed different thoughts to Smith — for example, the thought (a) that the tree *is to the left of* the house, (b) that the tree is *just* taller than the

roof of the house, (c) that there is only one tree in the yard, (d) that the tree is near the door, or (e) that there is not a cloud in the sky. In each of these cases, we would parse the image differently from our original parsing. For example, in the case of (a), the relation between the elements of the tree and house would be “_ is to the left of _,” and, in (b), (c), and (d), respectively, the roof, yard, and door would be elements.

Thus, we cannot determine what elements a thought has independently of determining what the thought is — once we grant the functional ambiguity of the structures we take to be identical with (or the causes of) thoughts. How we parse a thought (a bit of mental history) will depend on what we take that bit of mental history to be (in the context supplied). Since one and the same inner episode can serve an indefinite number of different functions, no one of which is privileged or preferred apart from a context, that episode can have an indefinite number of parsings, no one of which is privileged or preferred.⁷

But that fact is fatal to physicalistic theories like those of Pribram and Walker. To see why, we need only imagine Smith having mental image *m* and being in brain state *b* in several different contexts, contexts in which we would regard Smith as having different thoughts while he had one and the same image (e.g., contexts where Jones says to Smith that the tree is near the door, or to the left of the house). Now we have seen that Smith’s mental image may be parsed into different constituent elements depending on how it functions. But then we cannot parse *b* into constituent elements independently of determining what that function is.

To see this, suppose we were not sure whether the functionally ambiguous image of the tree and house correlated with Smith’s thought (a) that the tree is to the left of the house or (d) that the tree is near the door. How would we determine, then, what *b*’s structure was? We could not claim that there is one and only one way to parse *b*. That would be a concession that even the unique parsing of *b* was functionally ambiguous, and thus that the content or set of representational properties of *b* (or *m*, for that matter) was not a function of *b*’s structure after all. So that concession would be fatal to any theory purporting to analyze thoughts in terms of hardware descriptions of the brain. Presumably, then, we must say that we parse *b* differently depending on how we analyze its position in a sequence of events, or what we identify as the function of its corresponding mental image *m*. For example, one parsing of *b* would give us the brain structure for “_ is to the left of _,” and another, the brain structure for “_ is near _.” But then the structure of *b* cannot be something about Smith’s brain that simply

⁷ Analogously, in the case of perception, Jones may have been calling Smith’s attention to any of a number of different things while Smith looks at the house. Here, we might properly characterize Smith as perceiving different things, even though, phenomenologically, nothing need be different for Smith. So if Jones had said to Smith that there was not a cloud in the sky, this could be what Smith perceives, even though his mental image remains the same as in our original case. That is one reason why it is unacceptable to maintain that the elements of a perceptual experience can be determined independently of specifying a context in which the experience occurred, and thus why perceptual experiences are not merely constructs out of elements in a sensory manifold.

presents itself to us. We could not then assign *b* a structure independently of assigning it some role or position in a sequence of events or slice of life. Since a brain state does not determine its own function — since it may function differently in different contexts — we may properly assign to it more than one structure or parsing. Therefore, since brain states have no intrinsic function, they have no intrinsic or manifest structure.

Even apart from these considerations, the claim that brain states have a manifest structure is completely implausible. Any ordinary physical object may be parsed in different ways for different purposes. For example, a human hand may be considered to be composed of skin, bones, tendons and nails, or fingers, knuckles, palm and back, or perhaps cells of different sorts, or collections of atoms. And no one of those parsings is intrinsically preferable to the others. Similarly, a brain state may also be parsed in different ways — for example, into atoms, cells, neural nets, etc. — no one of which is intrinsically preferable to any other. So not only is it a mistake to suppose that the structure of a brain state can be determined independently of determining its function, it is also a mistake to suppose that a brain state has an inherent or privileged structure.

The idea that nature has a manifest structure takes many forms and appears not only in connection with theories of consciousness. I want, now, to examine its role in the so-called theory of *synchronicity*.

References

- Anderson, J., R.M. (1977). "A Holographic Model of Transpersonal Consciousness." *Journal of Transpersonal Psychology* 9: 119–128.
- Bursen, H.A. (1978). *Dismantling the Memory Machine*. Dordrecht, Boston, London: D. Reidel.
- Davidson, D. (1970). "Mental Events," In L. Foster & J. W. Swanson (eds.), *Experience and Theory*. Amherst: University of Massachusetts Press: 79–101.
- Dennett, D.C. (1978). *Brainstorms*. Montgomery, VT: Bradford Books.
- Goldberg, B. (1977). "A Problem with Anomalous Monism." *Philosophical Studies* 32: 175–180.
- Heil, J. (1978). "Traces of Things Past." *Philosophy of Science* 45: 60–67.
- Lurie, Y. (1973). *The Correspondence Thesis*. Doctoral Dissertation, Cornell University. (Ann Arbor: University Microfilms, 1977)
- Malcolm, N. (1977). *Memory and Mind*. Ithaca: Cornell University Press.
- Pribram, K. (1971). *Languages of the Brain*. Englewood Cliffs, NJ: Prentice Hall.
- Pribram, K. (1977). "Holonomy and Structure in the Organization of Perception." In U.M. Nicholas (Ed.), *Images, Perception and Knowledge*. Dordrecht: Reidel.
- Pribram, K., Nuwer, M., and Baron, R.U. (1974). "The Holographic Hypothesis of Memory Structure in Brain Function and Perception." In R.C. Atkinson, D.H. Krantz, R.C. Luce, & P. Suppes (eds.), *Contemporary Development in Mathematical Psychology*. San Francisco: Freeman.
- Walker, E. H. (1974). "Consciousness and Quantum Theory." In E.D. Mitchell (Ed.), *Psychic Exploration*. New York: Putnam: 544–568.
- Walker, E. H. (1975). "Foundations of Paraphysical and Parapsychological Phenomena." In L. Oteri (Ed.), *Quantum Physics and Parapsychology*. New York: Parapsychology Foundation: 1–53.
- Wittgenstein, L. (1955a). *Philosophical Investigations* (G.E.M. Anscombe, Trans.). G.E.M. Anscombe & R. R. (eds.). Oxford: Blackwell.
- Wittgenstein, L. (1955b). *Zettel* (G.E.M. Anscombe, Trans.). G.E.M. Anscombe & G.H. von Wright (eds.). Oxford: Blackwell.