

C. The Identity Theory

Is Consciousness a Brain Process?

U. T. Place

The thesis that consciousness is a process in the brain is put forward as a reasonable scientific hypothesis, not to be dismissed on logical grounds alone. The conditions under which two sets of observations are treated as observations of the same process, rather than as observations of two independent correlated processes, are discussed. It is suggested that we can identify consciousness with a given pattern of brain activity, if we can explain the subject's introspective observations by reference to the brain processes with which they are correlated. It is argued that the problem of providing a physiological explanation of introspective observations is made to seem more difficult than it really is by the "phenomenological fallacy," the mistaken idea that descriptions of the appearances of things are descriptions of the actual state of affairs in a mysterious internal environment.

I. Introduction

The view that there exists a separate class of events, mental events, which cannot be described in terms of the concepts employed by the physical sciences no longer commands the universal and unquestioning acceptance among philosophers and psychologists which it once did. Modern physicalism, however, unlike the materialism of the seventeenth and eighteenth centuries, is behavioristic. Consciousness on this view is either a special type of behavior, "sampling" or "running-back-and-forth" behavior as Tolman has it,¹ or a disposition to behave in a certain way, an itch, for example, being a temporary propensity to scratch. In the case of cognitive concepts like "knowing," "believing," "understanding," "remembering," and volitional concepts like "wanting" and "intending,"

there can be little doubt, I think, that an analysis in terms of dispositions to behave is fundamentally sound.² On the other hand, there would seem to be an intractable residue of concepts clustering around the notions of consciousness, experience, sensation, and mental imagery, where some sort of inner process story is unavoidable.³ It is possible, of course, that a satisfactory behavioristic account of this conceptual residuum will ultimately be found. For our present purposes, however, I shall assume that this cannot be done and that statements about pains and twinges, about how things look, sound, and feel, about things dreamed of or pictured in the mind's eye, are statements referring to events and processes which are in some sense private or internal to the individual of whom they are predicated. The question I wish to raise is whether in making this assumption we are inevitably committed to a dualist position in which sensations and mental images form a separate category of processes over and above the physical and physiological processes with which they are known to be correlated. I shall argue that an acceptance of inner processes does not entail dualism and that the thesis that consciousness is a process in the brain cannot be dismissed on logical grounds.

II. The "Is" of Definition and the "Is" of Composition

I want to stress from the outset that in defending the thesis that consciousness is a process in the brain, I am not trying to argue that when we describe our dreams, fantasies, and sensations we are talking about processes in our brains. That is, I am not claiming that statements about sensations and mental images are reducible to or

analyzable into statements about brain processes, in the way in which “cognition statements” are analyzable into statements about behavior. To say that statements about consciousness are statements about brain processes is manifestly false. This is shown (a) by the fact that you can describe your sensations and mental imagery without knowing anything about your brain processes or even that such things exist, (b) by the fact that statements about one’s consciousness and statements about one’s brain processes are verified in entirely different ways, and (c) by the fact that there is nothing self-contradictory about the statement “X has a pain but there is nothing going on in his brain.” What I do want to assert, however, is that the statement “Consciousness is a process in the brain,” although not necessarily true, is not necessarily false. “Consciousness is a process in the brain” in my view is neither self-contradictory nor self-evident; it is a reasonable scientific hypothesis, in the way that the statement “Lightning is a motion of electric charges” is a reasonable scientific hypothesis.

The all but universally accepted view that an assertion of identity between consciousness and brain processes can be ruled out on logical grounds alone derives, I suspect, from a failure to distinguish between what we may call the “is” of definition and the “is” of composition. The distinction I have in mind here is the difference between the function of the word “is” in statements like “A square is an equilateral rectangle,” “Red is a color,” “To understand an instruction is to be able to act appropriately under the appropriate circumstances,” and its function in statements like “His table is an old packing case,” “Her hat is a bundle of straw tied together with string,” “A cloud is a mass of water droplets or other particles in suspension.” These two types of “is” statements have one thing in common. In both cases it makes sense to add the qualification “and nothing else.” In this they differ from those statements in which the “is” is an “is” of predication; the statements “Toby is eighty years old and nothing else,” “Her hat is red and nothing else,” or “Giraffes are tall and nothing else,” for example, are nonsense. This logical feature may be described by saying that in both cases both the grammatical subject and the grammatical predicate are expressions which provide an adequate characterization of the state of affairs to which they both refer.

In another respect, however, the two groups of statements are strikingly different. State-

ments like “A square is an equilateral rectangle” are necessary statements which are true by definition. Statements like “His table is an old packing-case,” on the other hand, are contingent statements which have to be verified by observation. In the case of statements like “A square is an equilateral rectangle” or “Red is a color,” there is a relationship between the meaning of the expression forming the grammatical predicate and the meaning of the expression forming the grammatical subject, such that whenever the subject expression is applicable the predicate must also be applicable. If you can describe something as red then you must also be able to describe it as colored. In the case of statements like “His table is an old packing-case,” on the other hand, there is no such relationship between the meanings of the expressions “his table” and “old packing-case”; it merely so happens that in this case both expressions are applicable to and at the same time provide an adequate characterization of the same object. Those who contend that the statement “Consciousness is a brain process” is logically untenable, base their claim, I suspect, on the mistaken assumption that if the meanings of two statements or expressions are quite unconnected, they cannot both provide an adequate characterization of the same object or state of affairs: if something is a state of consciousness, it cannot be a brain process, since there is nothing self-contradictory in supposing that someone feels a pain when there is nothing happening inside his skull. By the same token we might be led to conclude that a table cannot be an old packing-case, since there is nothing self-contradictory in supposing that someone has a table, but is not in possession of an old packing-case.

III. The Logical Independence of Expressions and the Ontological Independence of Entities

There is, of course, an important difference between the table/packing-case and the consciousness/brain process case in that the statement “His table is an old packing-case” is a particular proposition which refers only to one particular case, whereas the statement “Consciousness is a process in the brain” is a general or universal proposition applying to all states of consciousness whatever. It is fairly clear, I think, that if we

lived in a world in which all tables without exception were packing-cases, the concepts of "table" and "packing-case" in our language would *not* have their present logically independent status. In such a world a table would be a species of packing-case in much the same way that red is a species of color. It seems to be a rule of language that whenever a given variety of object or state of affairs has two characteristics or sets of characteristics, one of which is unique to the variety of object or state of affairs in question, the expression used to refer to the characteristic or set of characteristics which defines the variety of object or state of affairs in question will always entail the expression used to refer to the other characteristic or set of characteristics. If this rule admitted of no exception it would follow that any expression which is logically independent of another expression which uniquely characterizes a given variety of object or state of affairs must refer to a characteristic or set of characteristics which is not normally or necessarily associated with the object or state of affairs in question. It is because this rule applies almost universally, I suggest, that we are normally justified in arguing from the logical independence of two expressions to the ontological independence of the states of affairs to which they refer. This would explain both the undoubted force of the argument that consciousness and brain processes must be independent entities because the expressions used to refer to them are logically independent and, in general, the curious phenomenon whereby questions about the furniture of the universe are often fought and not infrequently decided merely on a point of logic.

The argument from the logical independence of two expressions to the ontological independence of the entities to which they refer breaks down in the case of brain processes and consciousness, I believe, because this is one of a relatively small number of cases where the rule stated above does not apply. These exceptions are to be found, I suggest, in those cases where the operations which have to be performed in order to verify the presence of the two sets of characteristics inhering in the object or state of affairs in question can seldom if ever be performed simultaneously. A good example here is the case of the cloud and the mass of droplets or other particles in suspension. A cloud is a large semi-transparent mass with a fleecy texture suspended in the atmosphere whose shape is subject to continual and kaleidoscopic change. When observed at close quarters, however, it is found to consist of

a mass of tiny particles, usually water droplets, in continuous motion. On the basis of this second observation we conclude that a cloud is a mass of tiny particles and nothing else. But there is no logical connection in our language between a cloud and a mass of tiny particles; there is nothing self-contradictory in talking about a cloud which is not composed of tiny particles in suspension. There is no contradiction involved in supposing that clouds consist of a dense mass of fibrous tissue; indeed, such a consistency seems to be implied by many of the functions performed by clouds in fairy stories and mythology. It is clear from this that the terms "cloud" and "mass of tiny particles in suspension" mean quite different things. Yet we do not conclude from this that there must be two things, the mass of particles in suspension and the cloud. The reason for this, I suggest, is that although the characteristics of being a cloud and being a mass of tiny particles in suspension are invariably associated, we never make the observations necessary to verify the statement "That is a cloud" and those necessary to verify the statement "This is a mass of tiny particles in suspension" at one and the same time. We can observe the micro-structure of a cloud only when we are enveloped by it, a condition which effectively prevents us from observing those characteristics which from a distance lead us to describe it as a cloud. Indeed, so disparate are these two experiences that we use different words to describe them. That which is a cloud when we observe it from a distance becomes a fog or mist when we are enveloped by it.

IV. When Are Two Sets of Observations Observations of the Same Event?

The example of the cloud and the mass of tiny particles in suspension was chosen because it is one of the few cases of a general proposition involving what I have called the "is" of composition which does not involve us in scientific technicalities. It is useful because it brings out the connection between the ordinary everyday cases of the "is" of composition like the table/packing-case example and the more technical cases like "Lightning is a motion of electric charges" where the analogy with the consciousness/brain process case is most marked. The limitation of the cloud/tiny particles in suspension case is that it does not bring out sufficiently clearly the crucial problems of how the identity

of the states of affairs referred to by the two expressions is established. In the cloud case the fact that something is a cloud and the fact that something is a mass of tiny particles in suspension are both verified by the normal processes of visual observation. It is arguable, moreover, that the identity of the entities referred to by the two expressions is established by the continuity between the two sets of observations as the observer moves towards or away from the cloud. In the case of brain processes and consciousness there is no such continuity between the two sets of observations involved. A closer introspective scrutiny will never reveal the passage of nerve impulses over a thousand synapses in the way that a closer scrutiny of a cloud will reveal a mass of tiny particles in suspension. The operations required to verify statements about consciousness and statements about brain processes are fundamentally different.

To find a parallel for this feature we must examine other cases where an identity is asserted between something whose occurrence is verified by the ordinary processes of observation and something whose occurrence is established by special procedures. For this purpose I have chosen the case where we say that lightning is a motion of electric charges. As in the case of consciousness, however closely we scrutinize the lightning we shall never be able to observe the electric charges, and just as the operations for determining the nature of one's state of consciousness are radically different from those involved in determining the nature of one's brain processes, so the operations for determining the occurrence of lightning are radically different from those involved in determining the occurrence of a motion of electric charges. What is it, therefore, that leads us to say that the two sets of observations are observations of the same event? It cannot be merely the fact that the two sets of observations are systematically correlated such that whenever there is lightning there is always a motion of electric charges. There are innumerable cases of such correlations where we have no temptation to say that the two sets of observations are observations of the same event. There is a systematic correlation, for example, between the movement of the tides and the stages of the moon, but this does not lead us to say that records of tidal levels are records of the moon's stages or vice versa. We speak rather of a causal connection between two independent events or processes.

The answer here seems to be that we treat the two sets of observations as observations of the

same event in those cases where the technical scientific observations set in the context of the appropriate body of scientific theory provide an immediate explanation of the observations made by the man in the street. Thus we conclude that lightning is nothing more than a motion of electric charges, because we know that a motion of electric charges through the atmosphere, such as occurs when lightning is reported, gives rise to the type of visual stimulation which would lead an observer to report a flash of lightning. In the moon/tide case, on the other hand, there is no such direct causal connection between the stages of the moon and the observations made by the man who measures the height of the tide. The causal connection is between the moon and the tides, not between the moon and the measurement of the tides.

V. The Physiological Explanation of Introspection and the Phenomenological Fallacy

If this account is correct, it should follow that in order to establish the identity of consciousness and certain processes in the brain, it would be necessary to show that the introspective observations reported by the subject can be accounted for in terms of processes which are known to have occurred in his brain. In the light of this suggestion it is extremely interesting to find that when a physiologist, as distinct from a philosopher, finds it difficult to see how consciousness could be a process in the brain, what worries him is not any supposed self-contradiction involved in such an assumption, but the apparent impossibility of accounting for the reports given by the subject of his conscious processes in terms of the known properties of the central nervous system. Sir Charles Sherrington has posed the problem as follows:

The chain of events stretching from the sun's radiation entering the eye to, on the one hand, the contraction of the pupillary muscles, and on the other, to the electrical disturbances in the brain-cortex are all straightforward steps in a sequence of physical "causation," such as, thanks to science, are intelligible. But in the second serial chain there follows on, or attends, the stage of brain-cortex reaction an event or set of events quite inexplicable to us, which both as to themselves and as to the causal tie between them and what preceded them science does not help us; a set of events seemingly incommensurable with

any of the events leading up to it. The self "sees" the sun; it senses a two-dimensional disc of brightness, located in the "sky," this last a field of lesser brightness, and overhead shaped as a rather flattened dome, coping the self and a hundred other visual things as well. Of hint that this is within the head there is none. Vision is saturated with this strange property called "projection," the unargued inference that what it sees is at a "distance" from the seeing "self." Enough has been said to stress that in the sequence of events a step is reached where a physical situation in the brain leads to a psychical, which however contains no hint of the brain or any other bodily part . . . The supposition has to be, it would seem, two continuous series of events, one physico-chemical, the other psychical, and at times interaction between them.⁴

Just as the physiologist is not likely to be impressed by the philosopher's contention that there is some self-contradiction involved in supposing consciousness to be a brain process, so the philosopher is unlikely to be impressed by the considerations which lead Sherrington to conclude that there are two sets of events, one physico-chemical, the other psychical. Sherrington's argument, for all its emotional appeal, depends on a fairly simple logical mistake, which is unfortunately all too frequently made by psychologists and physiologists and not infrequently in the past by the philosophers themselves. This logical mistake, which I shall refer to as the "phenomenological fallacy," is the mistake of supposing that when the subject describes his experience, when he describes how things look, sound, smell, taste, or feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen, usually referred to in the modern psychological literature as the "phenomenal field." If we assume, for example, that when a subject reports a green after-image he is asserting the occurrence inside himself of an object which is literally green, it is clear that we have on our hands an entity for which there is no place in the world of physics. In the case of the green after-image there is no green object in the subject's environment corresponding to the description that he gives. Nor is there anything green in his brain; certainly there is nothing which could have emerged when he reported the appearance of the green after-image. Brain processes are not the sort of things to which color concepts can be properly applied.

The phenomenological fallacy on which this argument is based depends on the mistaken assumption that because our ability to describe

things in our environment depends on our consciousness of them, our descriptions of things are primarily descriptions of our conscious experience and only secondarily, indirectly, and inferentially descriptions of the objects and events in our environments. It is assumed that because we recognize things in our environment by their look, sound, smell, taste, and feel, we begin by describing their phenomenal properties, i.e. the properties of the looks, sounds, smells, tastes, and feels which they produce in us, and infer their real properties from their phenomenal properties. In fact, the reverse is the case. We begin by learning to recognize the real properties of things in our environment. We learn to recognize them, of course, by their look, sound, smell, taste, and feel; but this does not mean that we have to learn to describe the look, sound, smell, taste, and feel of things before we can describe the things themselves. Indeed, it is only after we have learned to describe the things in our environment that we learn to describe our consciousness of them. We describe our conscious experience not in terms of the mythological "phenomenal properties" which are supposed to inhere in the mythological "objects" in the mythological "phenomenal field," but by reference to the actual physical properties of the concrete physical objects, events, and processes which normally, though not perhaps in the present instance, give rise to the sort of conscious experience which we are trying to describe. In other words when we describe the after-image as green, we are not saying that there is something, the after-image, which is green; we are saying that we are having the sort of experience which we normally have when, and which we have learned to describe as, looking at a green patch of light.

Once we rid ourselves of the phenomenological fallacy we realize that the problem of explaining introspective observations in terms of brain processes is far from insuperable. We realize that there is nothing that the introspecting subject says about his conscious experiences which is inconsistent with anything the physiologist might want to say about the brain processes which cause him to describe the environment and his consciousness of that environment in the way he does. When the subject describes his experience by saying that a light which is in fact stationary appears to move, all the physiologist or physiological psychologist has to do in order to explain the subject's introspective observations is to show that the brain process which is causing the subject to describe his experience in

this way is the sort of process which normally occurs when he is observing an actual moving object and which therefore normally causes him to report the movement of an object in his environment. Once the mechanism whereby the individual describes what is going on in his environment has been worked out, all that is required to explain the individual's capacity to make introspective observations is an explanation of his

ability to discriminate between those cases where his normal habits of verbal descriptions are appropriate to the stimulus situation and those cases where they are not, and an explanation of how and why, in those cases where the appropriateness of his normal descriptive habits is in doubt, he learns to issue his ordinary descriptive protocols preceded by a qualificatory phrase like "it appears," "seems," "looks," "feels," etc.⁵

NOTES

1. E. C. Tolman, *Purposive Behaviour in Animals and Men* (Berkeley 1932).
2. L. Wittgenstein, *Philosophical Investigations* (Oxford 1953); G. Ryle, *The Concept of Mind* (1949).
3. Place, "The Concept of Heed," *British Journal of Psychology* XLV (1954), 243–55.
4. Sir Charles Sherrington, *The Integrative Action of the Nervous System* (Cambridge 1947), pp. xx–xxi.
5. I am greatly indebted to my fellow-participants in a

series of informal discussions on this topic which took place in the Department of Philosophy, University of Adelaide, in particular to Mr. C. B. Martin for his persistent and searching criticism of my earlier attempts to defend the thesis that consciousness is a brain process, to Professor D. A. T. Gasking, of the University of Melbourne, for clarifying many of the logical issues involved, and to Professor J. J. C. Smart for moral support and encouragement in what often seemed a lost cause.



Sensations and Brain Processes

J. J. C. Smart

Suppose that I report that I have at this moment a roundish, blurry-edged after-image which is yellowish towards its edge and is orange towards its centre. What is it that I am reporting?¹ One answer to this question might be that I am not reporting anything, that when I say that it looks to me as though there is a roundish yellowish orange patch of light on the wall I am expressing some sort of *temptation*, the temptation to say that there *is* a roundish yellowish orange patch on the wall (though I may know that there is not such a patch on the wall). This is perhaps Wittgenstein's view in the *Philosophical Investigations* (see paragraphs 367, 370). Similarly, when I "report" a pain, I am not really reporting anything (or, if you like, I am reporting in a queer sense of "reporting"), but am doing a sophisticated sort of wince. (See

paragraph 244: "The verbal expression of pain replaces crying and does not describe it." Nor does it describe anything else?)² I prefer most of the time to discuss an after-image rather than a pain, because the word "pain" brings in something which is irrelevant to my purpose: the notion of "distress." I think that "he is in pain" entails "he is in distress," that is, that he is in a certain agitation-condition.³ Similarly, to say "I am in pain" may be to do more than "replace pain behavior": it may be partly to report something, though this something is quite non-mysterious, being an agitation-condition, and so susceptible of behavioristic analysis. The suggestion I wish if possible to avoid is a different one, namely that "I am in pain" is a genuine report, and that what it reports is an irreducibly psychical something. And similarly the sugges-

tion I wish to resist is also that to say "I have a yellowish orange after-image" is to report something irreducibly psychical.

Why do I wish to resist this suggestion? Mainly because of Occam's razor. It seems to me that science is increasingly giving us a viewpoint whereby organisms are able to be seen as physico-chemical mechanisms:⁴ it seems that even the behavior of man himself will one day be explicable in mechanistic terms. There does seem to be, so far as science is concerned, nothing in the world but increasingly complex arrangements of physical constituents. All except for one place: in consciousness. That is, for a full description of what is going on in a man you would have to mention not only the physical processes in his tissue, glands, nervous system, and so forth, but also his states of consciousness: his visual, auditory, and tactual sensations, his aches and pains. That these should be *correlated* with brain processes does not help, for to say that they are *correlated* is to say that they are something "over and above." You cannot correlate something with itself. You correlate footprints with burglars, but not Bill Sikes the burglar with Bill Sikes the burglar. So sensations, states of consciousness, do seem to be the one sort of thing left outside the physicalist picture, and for various reasons I just cannot believe that this can be so. That everything should be explicable in terms of physics (together of course with descriptions of the ways in which the parts are put together—roughly, biology is to physics as radio-engineering is to electro-magnetism) except the occurrence of sensations seems to me to be frankly unbelievable. Such sensations would be "nomological danglers," to use Feigl's expression.⁵ It is not often realized how odd would be the laws whereby these nomological danglers would dangle. It is sometimes asked, "Why can't there be psychophysical laws which are of a novel sort, just as the laws of electricity and magnetism were novelties from the standpoint of Newtonian mechanics?" Certainly we are pretty sure in the future to come across new ultimate laws of a novel type, but I expect them to relate simple constituents: for example, whatever ultimate particles are then in vogue. I cannot believe that ultimate laws of nature could relate simple constituents to configurations consisting of perhaps billions of neurons (and goodness knows how many billion billions of ultimate particles) all put together for all the world as though their main purpose in life was to be a negative feed-

back mechanism of a complicated sort. Such ultimate laws would be like nothing so far known in science. They have a queer "smell" to them. I am just unable to believe in the nomological danglers themselves, or in the laws whereby they would dangle. If any philosophical arguments seemed to compel us to believe in such things, I would suspect a catch in the argument. In any case it is the object of this paper to show that there are no philosophical arguments which compel us to be dualists.

The above is largely a confession of faith, but it explains why I find Wittgenstein's position (as I construe it) so congenial. For on this view there are, in a sense, no sensations. A man is a vast arrangement of physical particles, but there are not, over and above this, sensations or states of consciousness. There are just behavioral facts about this vast mechanism, such as that it expresses a temptation (behavior disposition) to say "there is a yellowish-red patch on the wall" or that it goes through a sophisticated sort of wince, that is, says "I am in pain." Admittedly Wittgenstein says that though the sensation "is not a something," it is nevertheless "not a nothing either" (paragraph 304), but this need only mean that the word "ache" has a use. An ache is a thing, but only in the innocuous sense in which the plain man, in the first paragraph of Frege's *Foundations of Arithmetic*, answers the question "what is the number one?" by "a thing." It should be noted that when I assert that to say "I have a yellowish-orange after-image" is to express a temptation to assert the physical-object statement "there is a yellowish-orange patch on the wall," I mean that saying "I have a yellowish-orange after-image" is (partly) the exercise of the disposition⁶ which is the temptation. It is not to *report* that I have the temptation, any more than is "I love you" normally a report that I love someone. Saying "I love you" is just part of the behavior which is the exercise of the disposition of loving someone.

Though, for the reasons given above, I am very receptive to the above "expressive" account of sensation statements, I do not feel that it will quite do the trick. Maybe this is because I have not thought it out sufficiently, but it does seem to me as though, when a person says "I have an after-image," he *is* making a genuine report, and that when he says "I have a pain," he is doing more than "replace pain-behavior," and that "this more" is not just to say that he is in distress. I am not so sure, however, that to admit this is to admit that there are nonphysical corre-

lates of brain processes. Why should not sensations just be brain processes of a certain sort? There are, of course, well-known (as well as lesser-known) philosophical objections to the view that reports of sensations are reports of brain-processes, but I shall try to argue that these arguments are by no means as cogent as is commonly thought to be the case.

Let me first try to state more accurately the thesis that sensations are brain processes. It is not the thesis that, for example, “after-image” or “ache” means the same as “brain process of sort *X*” (where “*X*” is replaced by a description of a certain sort of brain process). It is that, in so far as “after-image” or “ache” is a report of a process, it is a report of a process that *happens to be* a brain process. It follows that the thesis does not claim that sensation statements can be *translated* into statements about brain processes.⁷ Nor does it claim that the logic of a sensation statement is the same as that of a brain-process statement. All it claims is that in so far as a sensation statement is a report of something, that something is in fact a brain process. Sensations are nothing over and above brain processes. Nations are nothing “over and above” citizens, but this does not prevent the logic of nation statements being very different from the logic of citizen statements, nor does it insure the translatability of nation statements into citizen statements. (I do not, however, wish to assert that the relation of sensation statements to brain-process statements is very like that of nation statements to citizen statements. Nations do not just *happen to be* nothing over and above citizens, for example. I bring in the “nations” example merely to make a negative point: that the fact that the logic of A-statements is different from that of B-statements does not insure that A’s are anything over and above B’s.)

Remarks on Identity

When I say that a sensation is a brain process or that lightning is an electric discharge, I am using “is” in the sense of strict identity. (Just as in the—in this case necessary—proposition “7 is identical with the smallest prime number greater than 5.”) When I say that a sensation is a brain process or that lightning is an electric discharge I do not mean just that the sensation is somehow spatially or temporally continuous with the brain process or that the lightning is just spatially or temporally continuous with the discharge. When on the other hand I say that the

successful general is the same person as the small boy who stole the apples I mean only that the successful general I see before me is a time slice⁸ of the same four-dimensional object of which the small boy stealing apples is an earlier time slice. However, the four-dimensional object which has the general-I-see-before-me for its late time slice is identical in the strict sense with the four-dimensional object which has the small-boy-stealing-apples for an early time slice. I distinguish these two senses of “is identical with” because I wish to make it clear that the brain-process doctrine asserts identity in the *strict* sense.

I shall now discuss various possible objections to the view that the processes reported in sensation statements are in fact processes in the brain. Most of us have met some of these objections in our first year as philosophy students. All the more reason to take a good look at them. Others of the objections will be more recondite and subtle.

Objection 1

Any illiterate peasant can talk perfectly well about his after-images, or how things look or feel to him, or about his aches and pains, and yet he may know nothing whatever about neurophysiology. A man may, like Aristotle, believe that the brain is an organ for cooling the body without any impairment of his ability to make true statements about his sensations. Hence the things we are talking about when we describe our sensations cannot be processes in the brain.

Reply

You might as well say that a nation of slug-abeds, who never saw the morning star or knew of its existence, or who had never thought of the expression “the Morning Star,” but who used the expression “the Evening Star” perfectly well, could not use this expression to refer to the same entity as we refer to (and describe as) “the Morning Star.”⁹

You may object that the Morning Star is in a sense not the very same thing as the Evening Star, but only something spatiotemporally continuous with it. That is, you may say that the Morning Star is not the Evening Star in the strict sense of “identity” that I distinguished earlier. I can perhaps forestall this objection by considering the slug-abeds to be New Zealanders and the

early risers to be Englishmen. Then the thing the New Zealanders describe as “the Morning Star” could be the very same thing (in the strict sense) as the Englishmen describe as “the Evening Star.” And yet they could be ignorant of this fact.

There is, however, a more plausible example. Consider lightning.¹⁰ Modern physical science tells us that lightning is a certain kind of electrical discharge due to ionization of clouds of water-vapor in the atmosphere. This, it is now believed, is what the true nature of lightning is. Note that there are not two things: a flash of lightning and an electrical discharge. There is one thing, a flash of lightning, which is described scientifically as an electrical discharge to the earth from a cloud of ionized water-molecules. The case is not at all like that of explaining a footprint by reference to a burglar. We say that what lightning really is, what its true nature as revealed by science is, is an electric discharge. (It is not the true nature of a footprint to be a burglar.)

To forestall irrelevant objections, I should like to make it clear that by “lightning” I mean the publicly observable physical object, lightning, not a visual sense-datum of lightning. I say that the publicly observable physical object lightning is in fact the electric discharge, not just a correlate of it. The sense-datum, or at least the having of the sense-datum, the “look” of lightning, may well in my view be a correlate of the electric discharge. For in my view it is a brain state *caused* by the lightning. But we should no more confuse sensations of lightning with lightning than we confuse sensations of a table with the table.

In short, the reply to Objection 1 is that there can be contingent statements of the form “A is identical with B,” and a person may well know that something is an A without knowing that it is a B. An illiterate peasant might well be able to talk about his sensations without knowing about his brain processes, just as he can talk about lightning though he knows nothing of electricity.

Objection 2

It is only a contingent fact (if it is a fact) that when we have a certain kind of sensation there is a certain kind of process in our brain. Indeed it is possible, though perhaps in the highest degree unlikely, that our present physiological theories will be as out of date as the ancient theory connecting mental processes with goings on

in the heart. It follows that when we report a sensation we are not reporting a brain-process.

Reply

The objection certainly proves that when we say “I have an after-image” we cannot *mean* something of the form “I have such and such a brain-process.” But this does not show that what we report (having an after-image) is not *in fact* a brain process. “I see lightning” does not *mean* “I see an electric discharge.” Indeed, it is logically possible (though highly unlikely) that the electrical discharge account of lightning might one day be given up. Again, “I see the Evening Star” does not *mean* the same as “I see the Morning Star,” and yet “the Evening Star and the Morning Star are one and the same thing” is a contingent proposition. Possibly Objection 2 derives some of its apparent strength from a “Fido”—Fido theory of meaning. If the meaning of an expression were what the expression named, then of course it *would* follow from the fact that “sensation” and “brain-process” have different meanings that they cannot name one and the same thing.

Objection 3¹¹

Even if Objections 1 and 2 do not prove that sensations are something over and above brain-processes, they do prove that the qualities of sensations are something over and above the qualities of brain-processes. That is, it may be possible to get out of asserting the existence of irreducibly psychic processes, but not out of asserting the existence of irreducibly psychic *properties*. For suppose we identify the Morning Star with the Evening Star. Then there must be some properties which logically imply that of being the Morning Star, and quite distinct properties which entail that of being the Evening Star. Again, there must be some properties (for example, that of being a yellow flash) which are logically distinct from those in the physicalist story.

Indeed, it might be thought that the objection succeeds at one jump. For consider the property of “being a yellow flash.” It might seem that this property lies inevitably outside the physicalist framework within which I am trying to work (either by “yellow” being an objective emergent property of physical objects, or else by being a power to produce yellow sense-data, where

“yellow,” in this second instantiation of the word, refers to a purely phenomenal or introspectible quality). I must therefore digress for a moment and indicate how I deal with secondary qualities. I shall concentrate on color.

First of all, let me introduce the concept of a normal percipient. One person is more a normal percipient than another if he can make color discriminations that the other cannot. For example, if A can pick a lettuce leaf out of a heap of cabbage leaves, whereas B cannot though he can pick a lettuce leaf out of a heap of beetroot leaves, then A is more normal than B. (I am assuming that A and B are not given time to distinguish the leaves by their slight difference in shape, and so forth.) From the concept of “more normal than” it is easy to see how we can introduce the concept of “normal.” Of course, Eskimos may make the finest discriminations at the blue end of the spectrum, Hottentots at the red end. In this case the concept of a normal percipient is a slightly idealized one, rather like that of “the mean sun” in astronomical chronology. There is no need to go into such subtleties now. I say that “This is red” means something roughly like “A normal percipient would not easily pick this out of a clump of geranium petals though he would pick it out of a clump of lettuce leaves.” Of course it does not exactly mean this: a person might know the meaning of “red” without knowing anything about geraniums, or even about normal percipients. But the point is that a person can be *trained* to say “This is red” of objects which would not easily be picked out of geranium petals by a normal percipient, and so on. (Note that even a color-blind person can reasonably assert that something is red, though of course he needs to use another human being, not just himself, as his “color meter.”) This account of secondary qualities explains their unimportance in physics. For obviously the discriminations and lack of discriminations made by a very complex neurophysiological mechanism are hardly likely to correspond to simple and nonarbitrary distinctions in nature.

I therefore elucidate colors as powers, in Locke’s sense, to evoke certain sorts of discriminatory responses in human beings. They are also, of course, powers to cause sensations in human beings (an account still nearer Locke’s). But these sensations, I am arguing, are identifiable with brain processes.

Now how do I get over the objection that a sensation can be identified with a brain process only if it has some phenomenal property, not

possessed by brain processes, whereby one-half of the identification may be, so to speak, pinned down?

My suggestion is as follows. When a person says, “I see a yellowish-orange after-image,” he is saying something like this: “*There is something going on which is like what is going on when I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange.*” (And there is no reason why a person should not say the same thing when he is having a veridical sense-datum, so long as we construe “like” in the last sentence in such a sense that something can be like itself.) Notice that the italicized words, namely “there is something going on which is like what is going on when,” are all quasi-logical or topic-neutral words. This explains why the ancient Greek peasant’s reports about his sensations can be neutral between dualistic metaphysics or my materialistic metaphysics. It explains how sensations can be brain-processes and yet how those who report them need know nothing about brain-processes. For he reports them only very abstractly as “something going on which is like what is going on when . . .” Similarly, a person may say “someone is in the room,” thus reporting truly that the doctor is in the room, even though he has never heard of doctors. (There are not two people in the room: “someone” *and* the doctor.) This account of sensation statements also explains the singular elusiveness of “raw feels”—why no one seems to be able to pin any properties on them.¹² Raw feels, in my view, are colorless for the very same reason that *something* is colorless. This does not mean that sensations do not have properties, for if they are brain-processes they certainly have properties. It only means that in speaking of them as being like or unlike one another we need not know or mention these properties.

This, then, is how I would reply to Objection 3. The strength of my reply depends on the possibility of our being able to report that one thing is like another without being able to state the respect in which it is like. I am not sure whether this is so or not, and that is why I regard Objection 3 as the strongest with which I have to deal.

Objection 4

The after-image is not in physical space. The brain-process is. So the after-image is not a brain-process.

Reply

This is an *ignoratio elenchi*. I am not arguing that the after-image is a brain-process, but that the experience of having an after-image is a brain-process. It is the *experience* which is reported in the introspective report. Similarly, if it is objected that the after-image is yellowy-orange but that a surgeon looking into your brain would see nothing yellowy-orange, my reply is that it is the experience of seeing yellowy-orange that is being described, and this experience is not a yellowy-orange something. So to say that a brain-process cannot be yellowy-orange is not to say that a brain-process cannot in fact be the experience of having a yellowy-orange after-image. There is, in a sense, no such thing as an after-image or a sense-datum, though there is such a thing as the experience of having an image, and this experience is described indirectly in material object language, not in phenomenal language, for there is no such thing.¹³ We describe the experience by saying, in effect, that it is like the experience we have when, for example, we really see a yellowy-orange patch on the wall. Trees and wallpaper can be green, but not the experience of seeing or imagining a tree or wallpaper. (Or if they are described as green or yellow this can only be in a derived sense.)

Objection 5

It would make sense to say of a molecular movement in the brain that it is swift or slow, straight or circular, but it makes no sense to say this of the experience of seeing something yellow.

Reply

So far we have not given sense to talk of experiences as swift or slow, straight or circular. But I am not claiming that “experience” and “brain-process” mean the same or even that they have the same logic. “Somebody” and “the doctor” do not have the same logic, but this does not lead us to suppose that talking about somebody telephoning is talking about someone over and above, say, the doctor. The ordinary man when he reports an experience is reporting that something is going on, but he leaves it open as to what sort of thing is going on, whether in a material solid medium, or perhaps in some sort of gaseous medium, or even perhaps in some sort of non-

spatial medium (if this makes sense). All that I am saying is that “experience” and “brain-process” may in fact refer to the same thing, and if so we may easily adopt a convention (which is not a change in our present rules for the use of experience words but an addition to them) whereby it would make sense to talk of an experience in terms appropriate to physical processes.

Objection 6

Sensations are private, brain processes are *public*. If I sincerely say, “I see a yellowish-orange after-image” and I am not making a verbal mistake, then I cannot be wrong. But I can be wrong about a brain-process. The scientist looking into my brain might be having an illusion. Moreover, it makes sense to say that two or more people are observing the same brain-process but not that two or more people are reporting the same inner experience.

Reply

This shows that the language of introspective reports has a different logic from the language of material processes. It is obvious that until the brain-process theory is much improved and widely accepted there will be no *criteria* for saying “Smith has an experience of such-and-such a sort” *except* Smith’s introspective reports. So we have adopted a rule of language that (normally) what Smith says goes.

Objection 7

I can imagine myself turned to stone and yet having images, aches, pains, and so on.

Reply

I can imagine that the electrical theory of lightning is false, that lightning is some sort of purely optical phenomenon. I can imagine that lightning is not an electrical discharge. I can imagine that the Evening Star is not the Morning Star. But it is. All the objection shows is that “experience” and “brain-process” do not have the same meaning. It does not show that an experience is not in fact a brain process.

This objection is perhaps much the same as one which can be summed up by the slogan:

“What can be composed of nothing cannot be composed of anything.”¹⁴ The argument goes as follows: on the brain-process thesis the identity between the brain-process and the experience is a contingent one. So it is logically possible that there should be no brain-process, and no process of any other sort, either (no heart process, no kidney process, no liver process). There would be the experience but no “corresponding” physiological process with which we might be able to identify it empirically.

I suspect that the objector is thinking of the experience as a ghostly entity. So it is composed of something, not of nothing, after all. On his view it is composed of ghost stuff, and on mine it is composed of brain stuff. Perhaps the counter-reply will be¹⁵ that the experience is simple and uncompounded, and so it is not composed of anything after all. This seems to be a quibble, for, if it were taken seriously, the remark “What can be composed of nothing cannot be composed of anything” could be recast as an a priori argument against Democritus and atomism and for Descartes and infinite divisibility. And it seems odd that a question of this sort could be settled a priori. We must therefore construe the word “composed” in a very weak sense, which would allow us to say that even an indivisible atom is composed of something (namely, itself). The dualist cannot really say that an experience can be composed of nothing. For he holds that experiences are something over and above material processes, that is, that they are a sort of ghost stuff. (Or perhaps ripples in an underlying ghost stuff.) I say that the dualist’s hypothesis is a perfectly intelligible one. But I say that experiences are not to be identified with ghost stuff but with brain stuff. This is another hypothesis, and in my view a very plausible one. The present argument cannot knock it down a priori.

Objection 8

The “beetle in the box” objection (see Wittgenstein, *Philosophical Investigations*, paragraph 293). How could descriptions of experiences, if these are genuine reports, get a foothold in language? For any rule of language must have public criteria for its correct application.

Reply

The change from describing how things are to describing how we feel is just a change from un-

inhibitedly saying “this is so” to saying “this looks so.” That is, when the naive person might be tempted to say, “There is a patch of light on the wall which moves whenever I move my eyes” or “A pin is being stuck into me,” we have learned how to resist this temptation and say “It *looks as though* there is a patch of light on the wallpaper” or “It *feels as though* someone were sticking a pin into me.” The introspective account tells us about the individual’s state of consciousness in the same way as does “I see a patch of light” or “I feel a pin being stuck into me”: it differs from the corresponding perception statement in so far as (a) in the perception statement the individual “goes beyond the evidence of his senses” in describing his environment and (b) in the introspective report he withholds descriptive epithets he is inclined to ascribe to the environment, perhaps because he suspects that they may not be appropriate to the actual state of affairs. Psychologically speaking, the change from talking about the environment to talking about one’s state of consciousness is simply a matter of inhibiting descriptive reactions not justified by appearances alone, and of disinhibiting descriptive reactions which are normally inhibited because the individual has learned that they are unlikely to provide a reliable guide to the state of the environment in the prevailing circumstances.¹⁶ To say that something looks green to me is to say that my experience is like the experience I get when I see something that really is green. In my reply to Objection 3, I pointed out the extreme openness or generality of statements which report experiences. This explains why there is no language of private qualities. (Just as “someone,” unlike “the doctor,” is a colorless word.)¹⁷

If it is asked what is the difference between those brain processes which, in my view, are experiences and those brain processes which are not, I can only reply that this is at present unknown. But it does not seem to me altogether fanciful to conjecture that the difference may in part be that between perception and reception (in Dr. D. M. MacKay’s terminology) and that the type of brain process which is an experience might be identifiable with MacKay’s active “matching response.”¹⁸

I have now considered a number of objections to the brain-process thesis. I wish now to conclude by some remarks on the logical status of the thesis itself. U. T. Place seems to hold that it is a straight-out scientific hypothesis.¹⁹ If so, he is partly right and partly wrong. If the issue is be-

tween (say) a brain-process thesis and a heart thesis, or a liver thesis, or a kidney thesis, then the issue is a purely empirical one, and the verdict is overwhelmingly in favor of the brain. The right sorts of things don't go on in the heart, liver, or kidney, nor do these organs possess the right sort of complexity of structure. On the other hand, if the issue is between a brain-or-heart-or-liver-or-kidney thesis (that is, some form of materialism) on the one hand and epiphenomenalism on the other hand, then the issue is not an empirical one. For there is no conceivable experiment which could decide between materialism and epiphenomenalism. This latter issue is not like the average straight-out empirical issue in science, but like the issue between the nineteenth-century English naturalist Philip Gosse²⁰ and the orthodox geologists and paleontologists of his day. According to Gosse, the earth was created about 4000 B.C. exactly as described in Genesis, with twisted rock strata, "evidence" of erosion, and so forth, and all sorts of fossils, all in their appropriate strata, just as if the usual evolutionist story had been true. Clearly this theory is in a sense irrefutable: no evidence can possibly tell against it. Let us ignore the theological setting in which Philip Gosse's hypothesis had been placed, thus ruling out objections of a theological kind, such as "what a queer God who would go to such elaborate lengths to deceive us." Let us suppose that it is held that the universe just *began* in 4004 B.C. with the initial conditions just every-

where as they were in 4004 B.C., and in particular that our own planet began with sediment in the rivers, eroded cliffs, fossils in the rocks, and so on. No scientist would ever entertain this as a serious hypothesis, consistent though it is with all possible evidence. The hypothesis offends against the principles of parsimony and simplicity. There would be far too many brute and inexplicable facts. Why are pterodactyl bones just as they are? No explanation in terms of the evolution of pterodactyls from earlier forms of life would any longer be possible. We would have millions of facts about the world as it was in 4004 B.C. that just have to be *accepted*.

The issue between the brain-process theory and epiphenomenalism seems to be of the above sort. (Assuming that a behavioristic reduction of introspective reports is not possible.) If it be agreed that there are no cogent philosophical arguments which force us into accepting dualism, and if the brain process theory and dualism are equally consistent with the facts, then the principles of parsimony and simplicity seem to me to decide overwhelmingly in favor of the brain-process theory. As I pointed out earlier, dualism involves a large number of irreducible psychophysical laws (whereby the "nomological danglers" dangle) of a queer sort, that just have to be taken on trust, and are just as difficult to swallow as the irreducible facts about the paleontology of the earth with which we are faced on Philip Gosse's theory.

NOTES

1. This paper takes its departure from arguments to be found in U. T. Place's "Is Consciousness a Brain Process?" (*British Journal of Psychology*, XLVII, 1956, 44-50). I have had the benefit of discussing Place's thesis in a good many universities in the United States and Australia, and I hope that the present paper answers objections to his thesis which Place has not considered, and presents his thesis in a more nearly unobjectionable form. This paper is meant also to supplement "The 'Mental' and the 'Physical,'" by H. Feigl (in *Minnesota Studies in the Philosophy of Science*, II, 370-497), which argues for much the same thesis as Place's.
2. Some philosophers of my acquaintance, who have the advantage over me in having known Wittgenstein, would say that this interpretation of him is too behavioristic. However, it seems to me a very natural interpretation of his printed words, and whether or not it is Wittgenstein's real view it is certainly an interesting and important one. I wish to consider it here as a possible rival both to the "brain-process" thesis and to straight-out old-fashioned dualism.
3. See Ryle, *Concept of Mind* (New York, 1949), p. 93.
4. On this point see Paul Oppenheim and Hilary Putnam, "Unity of Science as a Working Hypothesis," in *Minnesota Studies in the Philosophy of Science*, II, 3-36; also my note "Plausible Reasoning in Philosophy," *Mind*, LXVI (1957), 75-78.
5. Feigl, *op. cit.*, p. 428.
6. Wittgenstein did not like the word "disposition." I am using it to put in a nutshell (and perhaps inaccurately) the view which I am attributing to Wittgenstein. I should like to repeat that I do not wish to claim that my interpretation of Wittgenstein is correct. Some of those who knew him do not interpret him in this way. It is merely a view which I find myself extracting from his printed words and which I think is important and worth discussing for its own sake.
7. See Place, *op. cit.*, p. 45, near top, and Feigl, *op. cit.*, p. 390, near top.
8. See J. H. Woodger, *Theory Construction* (Chicago, 1939), p. 38 (International Encyclopedia of Unified Science, Vol. 2, No. 5). I here permit myself to speak loosely. For warnings against possible ways of going wrong with this sort of talk, see my note "Spatialising Time," *Mind*, LXIV (1955), 239-41.

9. Cf. Feigl, *op. cit.*, p. 439.
10. See Place, *op. cit.*, p. 47; also Feigl, *op. cit.* p. 438.
11. I think this objection was first put to me by Professor Max Black. I think it is the most subtle of any of those I have considered, and the one which I am least confident of having satisfactorily met.
12. See B. A. Farrell, "Experience," *Mind*, LIX (1950), especially 174.
13. Dr. J. R. Smythies claims that a sense-datum language could be taught independently of the material object language ("A Note on the Fallacy of the 'Phenomenological Fallacy,'" *British Journal of Psychology*, XLVIII, 1957, 141–144.) I am not so sure of this: there must be some public criteria for a person having got a rule wrong before we can teach him the rule. I suppose someone might *accidentally* learn color words by Dr. Smythies' procedure. I am not, of course, denying that we can learn a sense-datum language in the sense that we can learn to report our experience. Nor would Place deny it.
14. I owe this objection to Mr. C. B. Martin. I gather that he no longer wishes to maintain this objection, at any rate in its present form.
15. Martin did not make this reply, but one of his students did.
16. I owe this point to Place, in correspondence.
17. The "beetle in the box" objection is, *if it is sound*, an objection to *any* view, and in particular the Cartesian one, that introspective reports are genuine reports. So it is no objection to a weaker thesis that I would be concerned to uphold, namely, that if introspective reports of "experiences" are genuinely reports, then the things they are reports of are in fact brain processes.
18. See his article "Towards an Information-Flow Model of Human Behaviour," *British Journal of Psychology*, XLVII (1956), 30–43.
19. *Op. cit.*
20. See the entertaining account of Gosse's book *Omphalos* by Martin Gardner in *Fads and Fallacies in the Name of Science* (2nd ed., New York, 1957).

The "Mental" and the "Physical"

Herbert Feigl

E. Arguments Concerning the Identification of Sentience with Neural Events

I shall now present, as explicitly as I can, the reasons for an empirical identification of raw feels with neural processes. I shall also discuss several apparently trenchant arguments that have been advanced against this identity theory of the mental and the physical. It will be advisable first to state my thesis quite succinctly, and to elaborate the arguments for and against it afterwards.

Taking into consideration everything we have said so far about the scientific and the philosophical aspects of the mind–body problem, the following view suggests itself: The raw feels of direct experience as we "have" them, are empirically identifiable with the referents of certain specifiable concepts of molar behavior theory, and these in turn (this was argued in the preceding subsection *D*) are empirically identifiable

with the referents of some neurophysiological concepts. As we have pointed out, the word "mental" in present day psychology covers, however, not only the events and processes of direct experience (i.e., the raw feels), but also the unconscious events and processes, as well as the "intentional acts" of perception, introspective awareness, expectation, thought, belief, doubt, desire, volition, resolution, etc. I have argued above that since intentionality as such is to be analyzed on the one hand in terms of pure semantics (and thus falls under the category of the logical, rather than the psychological), it would be a category mistake of the most glaring sort to attempt a neurophysiological identification of this aspect of "mind." But since, on the other hand, intentional acts as occurrents in direct experience are introspectively or phenomenologically describable in something quite like raw-feel terms, a neural identification of this aspect of mind is *prima facie* not excluded on purely logical grounds. Unconscious processes, such as those described in psychoanalytic theory, are