In any case, it seems that the problem of the relation between the universe and the laboratory will be a knotty one to unravel, and perhaps it may replace the Thales problem as the central question in physics. Hopefully, it will take us less than 2500 years to solve it.

GERALD FEINBERG

PHYSICS DEPARTMENT
COLUMBIA UNIVERSITY

AN ARGUMENT FOR THE IDENTITY THEORY

I. INTRODUCTION

THE (Psychophysical) Identity Theory is the hypothesis that —not necessarily but as a matter of fact—every experience [1] is identical with some physical state.[2] Specifically, with some neurochemical state. I contend that we who accept the materialistic working hypothesis that physical phenomena have none but purely physical explanations must accept the identity theory. This is to say more than do most friends of the theory, who say only that we are free to accept it, and should for the sake of some sort of economy or elegance. I do not need to make a case for the identity theory on grounds of economy,[3] since I believe it can and should rest on a stronger foundation.

My argument is this: The definitive characteristic of any (sort of) experience as such is its causal role, its syndrome of most typical causes and effects. But we materialists believe that these causal roles which belong by analytic necessity to experiences belong in fact to certain physical states. Since those physical states possess the definitive characteristics of experience, they must be the experiences.

My argument parallels an argument which we will find uncontroversial. Consider cylindrical combination locks for bicycle chains. The definitive characteristic of their state of being unlocked is the causal role of that state, the syndrome of its most typical causes and effects: namely, that setting the combination typically causes the lock to be unlocked and that being unlocked

[1] Experiences herein are to be taken in general as universals, not as abstract particulars.

[2] States also are to be taken in general as universals. I shall not distinguish between processes, events, phenomena, and states in a strict sense.

[3] I am therefore invulnerable to Brandt's objection that the identity theory is not clearly more economical than a certain kind of dualism. "Doubts about the Identity Theory," in *Dimensions of Mind*, Sidney Hook, ed. (New York: NYU Press, 1960), pp. 57–67.

typically causes the lock to open when gently pulled. That is all we need know in order to ascribe to the lock the state of being or of not being unlocked. But we may learn that, as a matter of fact, the lock contains a row of slotted discs; setting the combination typically causes the slots to be aligned; and alignment of the slots typically causes the lock to open when gently pulled. So alignment of slots occupies precisely the causal role that we ascribed to being unlocked by analytic necessity, as the definitive characteristic of being unlocked (for these locks). Therefore alignment of slots is identical with being unlocked (for these locks). They are one and the same state.

II. The Nature of the Identity Theory

We must understand that the identity theory asserts that certain physical states are experiences, introspectible processes or activities, not that they are the supposed intentional objects that experiences are experiences *of*. If these objects of experience really exist separate from experiences of them, or even as abstract parts thereof, they may well also be something physical. Perhaps they are also neural, or perhaps they are abstract constituents of veridically perceived surroundings, or perhaps they are something else, or nothing at all; but that is another story. So I am not claiming that an experience of seeing red, say, is itself somehow a red neural state.

Shaffer has argued that the identity theory is impossible because (abstract particular) experiences are, by analytic necessity, unlocated, whereas the (abstract particular) neural events that they supposedly are have a location in part of the subject's nervous system.[4] But I see no reason to believe that the principle that experiences are unlocated enjoys any analytic, or other, necessity. Rather it is a metaphysical prejudice which has no claim to be respected. Or if there is, after all, a way in which it is analytic that experiences are unlocated, that way is irrelevant: perhaps in our presystematic thought we regard only concreta as located in a primary sense, and abstracta as located in a merely derivative sense by their inherence in located concreta. But this possible source of analytic unlocatedness for experiences does not meet the needs of Shaffer's argument. For neural events are abstracta too. Whatever unlocatedness accrues to experiences not because they are mental but because they are abstract must accrue as much to neural events. So it does not discriminate between the two.

[4] "Could Mental States Be Brain Processes?", this JOURNAL, **58**, 26 (Dec. 21, 1961): 813–822.

The identity theory says that experience-ascriptions have the same reference as certain neural-state-ascriptions: both alike refer to the neural states which are experiences. It does not say that these ascriptions have the same sense. They do not; experience-ascriptions refer to a state by specifying the causal role that belongs to it accidentally, in virtue of causal laws, whereas neural-state-ascriptions refer to a state by describing it in detail. Therefore the identity theory does not imply that whatever is true of experiences as such is likewise true of neural states as such, nor conversely. For a truth about things of any kind *as such* is about things of that kind not by themselves, but together with the sense of expressions by which they are referred to as things of that kind.[5] So it is pointless to exhibit various discrepancies between what is true of experiences as such and what is true of neural states as such. We can explain those discrepancies without denying psychophysical identity and without admitting that it is somehow identity of a defective sort.

We must not identify an experience itself with the attribute that is predicated of somebody by saying that he is having that experience. The former *is* whatever state it is that occupies a certain definitive causal role; the latter is the attribute of *being in* whatever state it is that occupies that causal role. By this distinction we can answer the objection that, since experience-ascriptions and neural-state-descriptions are admittedly never synonymous and since attributes are identical just in case they are predicated by synonymous expressions, therefore experiences and neural states cannot be identical attributes. The objection does establish a non-identity, but not between experiences and neural states. (It is unfair to blame the identity theory for needing the protection of so suspiciously subtle a distinction, for a parallel distinction is needed elsewhere. Blue is, for instance, the color of my socks, but blue is not the attribute predicated of things by saying they are the color of my socks, since '. . . is blue' and '. . . is the color of my socks' are not synonymous.)

III. The First Premise: Experiences Defined by Causal Roles

The first of my two premises for establishing the identity theory is the principle that the definitive characteristic of any experience as such is its causal role. The definitive causal role of an experi-

[5] Here I have of course merely applied to states Frege's doctrine of sense and reference. See "On Sense and Reference," in *Translations from the Philosophical Writings of Gottlob Frege*, Peter Geach and Max Black, eds. (New York: Oxford, 1960), pp. 56–78.

ence is expressible by a finite [6] set of conditions that specify its typical causes and its typical effects under various circumstances. By analytic necessity these conditions are true of the experience and jointly distinctive of it.

My first premise is an elaboration and generalization of Smart's theory that avowals of experience are, in effect, of the form 'What is going on in me is like what is going on in me when . . .' followed by specification of typical stimuli for, or responses to, the experience.[7] I wish to add explicitly that . . . may be an elaborate logical compound of clauses if necessary; that . . . must specify typical causes or effects of the experience, not mere accompaniments; that these typical causes and effects may include other experiences; and that the formula does not apply only to first-person reports of experience.

This is not a materialist principle, nor does it ascribe materialism to whoever speaks of experiences. Rather it is an account of the parlance common to all who believe that experiences are something or other real and that experiences are efficacious outside their own realm. It is neutral between theories—or a lack of any theory—about what sort of real and efficacious things experiences are: neural states or the like, pulsations of ectoplasm or the like, or just experiences and nothing else. It is not neutral, however, between all current theories of mind and body. Epiphenomenalist and parallelist dualism are ruled out as contradictory because they deny the efficacy of experience. Behaviorism as a thoroughgoing dispositional analysis of all mental states, including experiences,[8] is likewise ruled out as denying the reality and *a fortiori* the efficacy of experiences. For a pure disposition is a fictitious entity. The expressions that ostensibly denote dispositions are best construed as syncategorematic parts of statements of the lawlike regularities in which (as we say) the dispositions are manifest.

Yet the principle that experiences are defined by their causal

[6] It would do no harm to allow the set of conditions to be infinite, so long as it is recursive. But I doubt the need for this relaxation.

[7] *Philosophy and Scientific Realism* (New York: Humanities Press, 1963), ch. v. Smart's concession that his formula does not really translate avowals is unnecessary. It results from a bad example: 'I have a pain' is not translatable as 'What is going on in me is like what goes on when a pin is stuck into me', because the concept of pain might be introduced without mention of pins. Indeed; but the objection is no good against the translation 'What is going on in me is like what goes on when (i.e. when and because) my skin is damaged'.

[8] Any theory of mind and body is compatible with a dispositional analysis of mental states other than experiences or with so-called "methodological behaviorism."

roles is itself behaviorist in origin, in that it inherits the behaviorist discovery that the (ostensibly) causal connections between an experience and its typical occasions and manifestations somehow contain a component of analytic necessity. But my principle improves on the original behavioristic embodiment of that discovery in several ways:

First, it allows experiences to be something real and so to be the effects of their occasions and the causes of their manifestations, as common opinion supposes them to be.

Second, it allows us to include other experiences among the typical causes and effects by which an experience is defined. It is crucial that we should be able to do so in order that we may do justice, in defining experiences by their causal roles, to the introspective accessibility which is such an important feature of any experience. For the introspective accessibility of an experience is its propensity reliably to cause other (future or simultaneous) experiences directed intentionally upon it, wherein we are aware of it. The requisite freedom to interdefine experiences is not available in general under behaviorism; interdefinition of experiences is permissible only if it can in principle be eliminated, which is so only if it happens to be possible to arrange experiences in a hierarchy of definitional priority. We, on the other hand, may allow interdefinition with no such constraint. We may expect to get mutually interdefined families of experiences, but they will do us no harm. There will be no reason to identify anything with one experience in such a family without regard to the others —but why should there be? Whatever occupies the definitive causal role of an experience in such a family does so by virtue of its own membership in a causal isomorph of the family of experiences, that is, in a system of states having the same pattern of causal connections with one another and the same causal connections with states outside the family, viz., stimuli and behavior. The isomorphism guarantees that if the family is identified *throughout* with its isomorph then the experiences in the family will have their definitive causal roles. So, *ipso facto,* the isomorphism requires us to accept the identity of all the experiences of the family with their counterparts in the causal isomorph of the family.[9]

[9] Putnam discusses an analogous case for machines: a family of ("logical" or "functional") states defined by their causal roles and mutually interdefined, and a causally isomorphic system of ("structural") states otherwise defined. He does not equate the correlated logical and structural states. "Minds and Machines," in *Dimensions of Mind,* pp. 148–179.

Third, we are not obliged to define an experience by the causes and effects of exactly all and only its occurrences. We can be content rather merely to identify the experience as that state which is *typically* caused in thus-and-such ways and *typically* causes thus-and-such effects, saying nothing about its causes and effects in a (small) residue of exceptional cases. A definition by causes and effects in typical cases suffices to determine what the experience is, and the fact that the experience has some characteristics or other besides its definitive causal role confers a sense upon ascriptions of it in some exceptional cases for which its definitive typical causes and effects are absent (and likewise upon denials of it in some cases for which they are present). Behaviorism does not acknowledge the fact that the experience is something apart from its definitive occasions and manifestations, and so must require that the experience be defined by a strictly necessary and sufficient condition in terms of them. Otherwise the behaviorist has merely a partial explication of the experience by criteria, which can never give more than a presumption that the experience is present or absent, no matter how much we know about the subject's behavior and any lawlike regularities that may govern it. Relaxation of the requirement for a strictly necessary and sufficient condition is welcome. As anybody who has tried to implement behaviorism knows, it is usually easy to find conditions which are *almost* necessary and sufficient for an experience. All the work—and all the complexity which renders it incredible that the conditions found should be known implicitly by every speaker—comes in trying to cover a few exceptional cases. In fact, it is just impossible to cover some atypical cases of experiences behavioristically: the case of a perfect actor pretending to have an experience he does not really have; and the case of a total paralytic who cannot manifest any experience he does have (both cases under the stipulation that the pretense or paralysis will last for the rest of the subject's life no matter what happens, in virtue of regularities just as lawlike as those by which the behaviorist seeks to define experiences).

It is possible, and probably good analytic strategy, to reconstrue any supposed pure dispositional state rather as a state defined by its causal role. The advantages in general are those we have seen in this case: the state becomes recognized as real and efficacious; unrestricted mutual interdefinition of the state and others of its sort becomes permissible; and it becomes intelligible

that the state may sometimes occur despite prevention of its definitive manifestations.[10]

I do not offer to prove my principle that the definitive characteristics of experiences as such are their causal roles. It would be verified by exhibition of many suitable analytic statements saying that various experiences typically have thus-and-such causes and effects. Many of these statements have been collected by behaviorists; I inherit these although I explain their status somewhat differently. Behaviorism is widely accepted. I am content to rest my case on the argument that my principle can accommodate what is true in behaviorism and can escape attendant difficulties.

IV. The Second Premise: Explanatory Adequacy of Physics

My second premise is the plausible hypothesis that there is some unified body of scientific theories, of the sort we now accept, which together provide a true and exhaustive account of all physical phenomena (i.e. all phenomena describable in physical terms). They are unified in that they are cumulative: the theory governing any physical phenomenon is explained by theories governing phenomena out of which that phenomenon is composed and by the way it is composed out of them. The same is true of the latter phenomena, and so on down to fundamental particles or fields governed by a few simple laws, more or less as conceived of in present-day theoretical physics. I rely on Oppenheim and Putnam for a detailed exposition of the hypothesis that we may hope to find such a unified physicalistic body of scientific theory and for a presentation of evidence that the hypothesis is credible.[11]

A confidence in the explanatory adequacy of physics is a vital part, but not the whole, of any full-blooded materialism. It is the empirical foundation on which materialism builds its superstructure of ontological and cosmological doctrines, among them the identity theory. It is also a traditional and definitive working hypothesis of natural science—what scientists say nowadays to the contrary is defeatism or philosophy. I argue that whoever shares this confidence must accept the identity theory.

My second premise does not rule out the existence of non-

[10] Quine advocates this treatment of such dispositional states as are worth saving in *Word and Object* (Cambridge, Mass.: MIT Press, and New York: Wyley, 1960), pp. 222–225. "They are conceived as built-in, enduring structural traits."

[11] "Unity of Science as a Working Hypothesis," in *Minnesota Studies in the Philosophy of Science*, II, Herbert Feigl, Michael Scriven, and Grover Maxwell, eds. (Minneapolis: Univ. of Minnesota Press, 1958), pp. 3–36.

physical phenomena; it is not an ontological thesis in its own right. It only denies that we need ever explain physical phenonmena by nonphysical ones. Physical phenomena are physically explicable, or they are utterly inexplicable insofar as they depend upon chance in a physically explicable way, or they are methodologically acceptable primitives. All manner of nonphysical phenomena may coexist with them, even to the extent of sharing the same space-time, provided only that the nonphysical phenomena are entirely inefficacious with respect to the physical phenomena. These coexistent nonphysical phenomena may be quite unrelated to physical phenomena; they may be causally independent but for some reason perfectly correlated with some physical phenomena (as experiences are, according to parallelism); they may be epiphenomena, caused by some physical phenomena but not themselves causing any (as experiences are, according to epiphenomenalism). If they are epiphenomena they may even be correlated with some physical phenomena, perfectly and by virtue of a causal law.

V. Conclusion of the Argument

But none of these permissible nonphysical phenomena can be experiences. For they must be entirely inefficacious with respect to all physical phenomena. But all the behavioral manifestations of experiences are (or involve) physical phenomena and so cannot be effects of anything that is inefficacious with respect to physical phenomena. These behavioral manifestations are among the typical effects definitive of any experience, according to the first premise. So nothing can be an experience that is inefficacious with respect to physical phenomena. So nothing can be an experience that is a nonphysical phenomenon of the sort permissible under the second premise. From the two premises it follows that experiences are some physical phenomena or other.

And there is little doubt which physical phenomena they must be. We are far from establishing positively that neural states occupy the definitive causal roles of experiences, but we have no notion of any other physical phenomena that could possibly occupy them, consistent with what we do know. So if nonphysical phenomena are ruled out by our confidence in physical explanation, only neural states are left. If it could be shown that neural states do not occupy the proper causal roles, we would be hard put to save materialism itself.

A version of epiphenomenalism might seem to evade my argument: let experiences be nonphysical epiphenomena, precisely cor-

related according to a causal law with some simultaneous physical states which are themselves physically (if at all) explicable. The correlation law (it is claimed) renders the experiences and their physical correlates causally equivalent. So the nonphysical experiences have their definitive physical effects after all—although they are not needed to explain those effects, so there is no violation of my second premise (since the nonphysical experiences redundantly redetermine the effects of their physical correlates). I answer thus: at best, this position yields nonphysical experiences alongside the physical experiences, duplicating them, which is not what its advocates intend. Moreover, it is false that such a physical state and its epiphenomenal correlate are causally equivalent. The position exploits a flaw in the standard regularity theory of cause. We know on other grounds that the theory must be corrected to discriminate between genuine causes and the spurious causes which are their epiphenomenal correlates. (The "power on" light does not cause the motor to go, even if it is a lawfully perfect correlate of the electric current that really causes the motor to go.) Given a satisfactory correction, the nonphysical correlate will be evicted from its spurious causal role and thereby lose its status as the experience. So this epiphenomenalism is not a counterexample.

The dualism of the common man holds that experiences are nonphysical phenomena which are the causes of a familiar syndrome of physical as well as nonphysical effects. This dualism is a worthy opponent, daring to face empirical refutation, and in due time it will be rendered incredible by the continuing advance of physicalistic explanation. I have been concerned to prevent dualism from finding a safe fall-back position in the doctrine that experiences are nonphysical and physically inefficacious. It is true that such phenomena can never be refuted by any amount of scientific theory and evidence. The trouble with them is rather that they cannot be what we call experiences. They can only be the non-physical epiphenomena or correlates of physical state which are experiences. If they are not the experiences themselves, they cannot rescue dualism when it is hard-pressed. And if they cannot do that, nobody has any motive for believing in them. Such things may be—but they are of no consequence.

DAVID K. LEWIS

HARVARD UNIVERSITY