

Here the question is clearly of the same logical nature as the queries: "Do butterflies feel?" "Do fishworms, when put on the hook, feel pain?" "Do plants have feelings?" "Do human embryos, four months old, have any direct experience?" I shall try to clarify the nature of these questions in the following subsection. For the moment it must suffice to point out that here we have to do with the distinction between "mental" (in the sense of *sentience*) and *physical₁*. *Intelligence*, in contrast to *sentience*, is clearly definable in *physical₁* terms. But as to whether *sentience* is so definable is perhaps the central perplexity among the mind-body puzzles.

But now to complete our analysis of the meanings of "physical": We have distinguished "physical₁" and "physical₂". By "physical₁ terms" I mean *all* (empirical) terms whose specification of meaning essentially involves logical (necessary or, more usually, probabilistic) connections with the intersubjective observation language, as well as the terms of this observation language itself. Theoretical concepts in physics, biology, psychology, and the social sciences hence are all—at least—*physical₁* concepts. By "physical₂" I mean the kind of theoretical concepts (and statements) which are sufficient for the *explanation*, i.e., the deductive or probabilistic derivation, of the observation statements regarding the inorganic (lifeless) domain of nature. If my conjecture (discussed above) is correct, then the scopes of *theoretical* "physical₁" and "physical₂" terms are the same. *If*, however, there is genuine emergence, i.e., logical underivability, in the domains of organic, mental, and/or social phenomena, then the scope of "physical₂" terms is clearly narrower than that of "physical₁" theoretical terms.

Within the category of "physical₁" terms, it is clearly important to distinguish observation terms from theoretical terms; and among the latter several levels may methodologically, if not logically, be distinguished. For example, the concepts of classical thermodynamics form one level, and the concepts of statistical or molecular mechanics (in terms of which those of thermodynamics, with certain modifications, appealed to James' own principle of pragmatism (derived from Peirce's meaning criterion which anticipated the essentially equivalent later operationist and logical-positivist formulations of the criterion). But E. A. Singer in turn was incisively criticized by D. S. Miller (224), who many years later (226) attacked on the same grounds the much more subtle linguistic behaviorism expounded in Gilbert Ryle's *The Concept of Mind*.

can be defined) form a "higher" level. The concepts of molar behavior theory are related analogously to those of the higher level of neurophysiology; and so on *mutatis mutandis*, throughout the various fields of scientific theories.

We conclude that to say "x is physical" is highly ambiguous. There is first the obvious distinction between the physical languages (physical language designators) and physical objects (physical language designata). This distinction carries through the two further distinctions and does not, for our purposes, require elaborate discussion. To illustrate, an electromagnetic field, just as the planet Jupiter, are designata of physical language terms. However, the observation terms of the physical₁ language serve also as the evidential basis of the physical₁ or physical₂ theoretical languages. *Theoretical* terms are here conceived as not explicitly definable on the basis of observation terms (cf. Carnap, 73; Feigl, 110; Sellars, 315), but as specified by postulates and by correspondence rules relating them to the terms of the observation language. And, to restate this in different words, if there is no genuine emergence in the logical sense above the level of lifeless phenomena, then there is no basic distinction between the theoretical terms of the *physical₁* and *physical₂* languages. That is to say that the theoretical terms of biology and psychology are explicitly definable on the basis of the theoretical concepts of physics in the same sense as the theoretical terms of chemistry (e.g., the chemical bond) are nowadays explicitly definable on the basis of the theoretical terms of the *physical₂* language (i.e., of the atomic and quantum theories).

The central questions of the mind-body problem then come down to this: are the concepts of introspective psychology—relating to phenomenal data or phenomenal fields—definable on the basis of physical₁ theoretical terms, and if so, are they also definable on the basis of physical₂ (theoretical) terms? The first question is a matter for philosophical analysis. The second question is, at the present level of scientific research, undecided, though my personal (admittedly bold and risky) guess is that future scientific progress will decide it affirmatively. We turn now to a discussion of the first question primarily, but occasional remarks about the second question will also be ventured.

B. *The Inference to Other Minds*. Behaviorism and phenomenalism display interesting similarities as well as fundamental differences. According to logical behaviorism, the concepts of mental states, disposi-

tions, and events are logical constructions based on (physical₁) characterizations of behavior. According to the more recent formulations of physicalism (Feigl, 113, 116; Carnap, 73; Sellars, 315) the "logical construction" thesis is inadequate and has to be replaced by an analysis in terms of postulates and correspondence rules. Very simply and very roughly, this means—in the material mode of speech—that for physicalism mental states are inferential ("illata," cf. Reichenbach, 273). Contrariwise, modern phenomenism (Carnap, 60; Ayer, 12; Goodman, 135) had maintained that the concepts of physical things, states, dispositions, and events are logical constructions based on concepts designating the phenomena of immediate experience. And in the "revised" version of phenomenism, i.e., a genuinely realistic epistemology based on phenomenal data, a doctrine which should not be called "phenomenalism" at all, the concepts of physical objects are inferential ("hypothetical constructs," "illata"). But this doctrine is in many of its tenets consonant with classical critical realism (von Hartmann, Külpe, Schlick, R. W. Sellars, D. Drake, C. A. Strong, J. B. Pratt, A. O. Lovejoy, G. Santayana). In contradistinction to critical realism, there is the earlier doctrine of neutral monism developed by the neorealists, especially E. B. Holt and Bertrand Russell (before his later critical realism), and historically rooted in the positivism and empiriocriticism of Hume, Mill, Mach, and Avenarius. Russell (284, 287) was the primary influence in Carnap's early epistemology (60, 61); and this sort of neutral monism was also adopted in prefatory philosophical remarks of some psychologists like E. C. Tolman (336), C. C. Pratt (260), and others.

The distinctive mark of neutral monism is a conception of the "given" which is (1) subjectless, i.e., it does not allow for the use of the personal pronoun "I"; and (2) is "neutral" in the sense that the given is characterizable as neither "mental" nor "physical." It maintains that both mentalistic concepts (the concepts of psychology) and physical concepts (those of physics) are logically constituted out of the more basic concepts designating neutral data. Psychology and physics are here understood as more or less systematic knowledge both on the level of common life, and on the more advanced level of science. Disregarding some technical logical questions, the data upon which the construction is based turn out to be items of immediate experience (sentience) and are thus "mental" after all, in one of the two senses of "mental" which we have been at pains to explicate.

This is not the place to review the many arguments * which have been advanced in the refutation of phenomenism. If an epistemology with a phenomenal basis can at all be worked out satisfactorily, then these data have to be conceived as lawfully related to the physical objects of everyday life. This means that the doctrine of logical constructionism or reductionism, i.e., of the explicit definability of physical concepts in terms of phenomenal concepts, has to be abandoned. The logical relations involved here are synthetic, and the translatability thesis is not just utopian (owing to the always admitted complexities), but completely inadequate, if not quixotic. I remain unimpressed with the significance of Craig's theorem (cf. Hempel's essay in the present volume) in this connection. An infinite set of postulates is not what phenomenists ever had in mind. And I believe there are other grave objections to that sort of a translatability doctrine. The kind of translatability which Craig's theorem allows for concerns only the empirical content of theories in the sense of all conceivable evidential (confirming) statements, but not in the sense of the factual reference of the postulates (and, hence, of the theorems).

Mutatis mutandis, it is now realized in many philosophical and psychological quarters † that the thesis of the translatability of statements about mental states (in phenomenal language) into statements about peripheral behavior (in descriptive, not theoretical physical₁ language) must also be repudiated.

With this firmly established orientation, the inference of sentience (raw feels) in other organisms seems *prima facie* restored to its original form as an argument from analogy. I have no doubt that analogy is the essential criterion for the ascription of sentience. But a closer look at the logic of the inference will prove worthwhile. The inference from peripheral behavior to central processes, very much like the inference from skulls to brains contained in them, is intersubjectively confirmable, and this in the sense that independent intersubjective evidence for the truth of these conclusions is in principle available. Just this is, of course,

* Cf. Freytag (128); Külpe (191); Broad (50, 51); Schlick (298); Reichenbach (273); Pap (248); Lovejoy (204); R. W. Sellars (307); W. S. Sellars (308); B. Russell (288); Kneale (179); Beck (24); Feigl (110, 111); Berlin (35); Watling (342); Braithwaite (48); E. J. Nelson (237, 238); and now, after a drastic change in outlook, even Ayer (18) is close to a critical realist position.

† Cf. Hempel (146); Carnap (64, 67, 73); Kaufmann (175); Jacobs (163); Pap (242, 243, 245, 248); Ayer (18); Feigl (113, 116); Cronbach and Meehl (79); Scriven (306).

not the case for the conclusions regarding mental states, if by mental states (*sentience, raw feels*) one means something that is not identifiable (i.e., not explicitly definable in physical₁ terms) with either overt-behavioral or central-neural states or processes.* If, contrary to the suggested orientation, such identifications could be made, i.e., if explicit definition could plausibly be given as an analysis of the meaning of phenomenal terms, then indeed no analogical inferences would be required. Nevertheless, considerations of analogy would be suggestive, though never decisive, for the terminological conventions according to which we apply or refrain from applying phenomenal terms to the behavior of animals and plants (let alone lifeless things).

If, however, phenomenal terms are logically irreducible to physicalistic terms, then parallelistic (epiphenomenalist) dualism is the most plausible alternative view. But interactionistic dualism is empirically much less defensible, and its methodological orientation too defeatist, to be acceptable to the current scientific outlook (cf. section II, above). And epiphenomenalism also has generally been considered objectionable because it denies the causal efficacy of raw feels; and because it introduces peculiar lawlike relations between cerebral events and mental events. These correlation laws are utterly different from any other laws of (physical₂) science in that, first, they are nomological "danglers," i.e., relations which connect intersubjectively confirmable events with events which *ex hypothesi* are in principle not intersubjectively and independently confirmable. Hence, the presence or absence of phenomenal data is not a difference that could conceivably make a difference in the confirmatory physical₁-observational evidence, i.e., in the publicly observable behavior, or for that matter in the neural processes observed or inferred by the neurophysiologists. And second, these correlation laws would, unlike other correlation laws in the natural sciences, be (again *ex hypothesi*) absolutely underivable from the premises of even the most inclusive and enriched set of postulates of any future theoretical physics or biology.

No wonder then that after a period of acquiescence with epiphenomenalism during the last century (T. H. Huxley, et al.), the behaviorist

* This is my way of stating succinctly the puzzle of "Other Minds" as it is understood in the long (unfinished) sequence of agonizing articles by John Wisdom (354), and in many other authors' publications, notably: Carnap (61, 62); Schlick (299); Ayer (15, 18); Austin (10); Pap (243, 248); Hampshire (141); Watling (341); Mellor (223).

movement in psychology took hold, and exercised an unprecedented influence in so many quarters. Behaviorists, in their way, repressed the problem in that they either denied the existence of raw feels (materialism); or in that they defined them in physical₁-observation terms (logical behaviorism); or they maintained that the subject matter of scientific and experimental psychology can be nothing but behavior (methodological behaviorism), which leaves the existence of raw feels an open question, but as of no relevance to science. Our previous discussions have, I trust, clearly indicated that behaviorism in the first sense is absurdly false; in the second sense it is inadequate as a logical analysis of the meaning of phenomenal terms; and in the third sense, it is an admittedly fruitful but limited program of research, but it entails no conclusion directly relevant to the central philosophical issue.

The repudiation of radical behaviorism and of logical behaviorism entails the acceptance of some sort of parallelistic doctrine. Recent arguments for this position * are *prima facie* highly persuasive. The basic point is simply that each of us knows his own states of immediate experience by acquaintance, and that by analogical reasoning we can infer similar, though never directly inspectable, states of experience in others. Direct inspection of the mental states of others is now generally considered a logical impossibility. For example, the subjunctive conditional, "If I were you, I would experience your pain," is not merely counterfactual, but counterlogical in that the antecedent of the conditional involves an outright inconsistency. The air of plausibility of the mentioned subjunctive conditional derives from entirely other, quite legitimate types of subjunctive conditionals, such as "If I had a broken leg (as you do), I should feel pain"; or "If I had (some traits of) your personality, flattery would please me." The logical grammar of personal proper names (or pronouns) however is such that it is downright self-contradictory to say (in a reasonably constructed and interpreted language) that Smith is Jones, or that I am you. The Mont Blanc cannot conceivably be identical with Mt. Everest!

Indirect verification or confirmation of statements regarding the mental states of other persons is however clearly possible once we have established laws regarding the correlation of the Φ 's with the ψ 's for our own case. And as we have pointed out, these laws could in prin-

* By Pap (243); Hampshire (141); Watling (341); Ayer (18).

ciple be most directly established with the help of an autocerebroscope. On the level of common life, of course, the correlations between neural and mental states are totally unknown. But a great many behavioral indicators are constantly being used in the (probabilistic) ascription of mental states. Logical analysis (Carnap, 73; Scriven, 306; Feyerabend, 119; Watling, 342; Feigl, 110, 111, 112, 114) has, I think, quite convincingly demonstrated the need for distinguishing the evidential bases from the factual reference of concepts and statements. The behavioral indicators serve as evidential bases for the ascription of mental states. Only the person who experiences the mental state can directly verify its occurrence. But there is no reason whatever to assume that when A reports his mental state, and B talks about it on the basis of behavioral evidence (or, if this is feasible, on the basis of neurophysiological evidence), that what they are talking about is not the very same mental state. This is indeed the way in which ordinary communication is understood. For example, if the doctor tells me a moment before lancing my abscess, "This will hurt," it is I who can directly verify this prediction. Moreover, most of us have learned from childhood on how to conceal our thoughts, feelings, sentiments, how to dissimulate, play-act, etc. And so we can justifiably say that behavioral symptoms do not reliably indicate mental states. In the light of the basic principles of normal induction and analogy, involving symmetry considerations, solipsism (with its arbitrary asymmetries) must be regarded as an absurdly false, rather than as a meaningless doctrine.

If we had completely adequate and detailed knowledge of the neural processes in human brains, and the knowledge of the one-one, or at least one-many ψ - Φ correlation laws, then a description of a neural state would be completely reliable evidence (or a genuine criterion) for the occurrence of the corresponding mental state. If these central neural events are essential intermediate links in the causal chain which connects stimuli with responses, then these central states are (probabilistically) inferable from stimulus-response situations. In this respect they have a logical status similar to the mental states as they are inferred from behavior in everyday life, or as the basis of psychological test situations. One may therefore wonder whether two steps of inference are really needed for a full logical reconstruction of the scientific ascription of mental states to other persons; the first step being the one from overt behavior to central neural events, and the second step being the one

from neural events to mental states. I shall return to this question in subsection E, where I shall discuss the arguments for and against the identification of raw feels with the denotata of certain theoretical physical₁ (or physical₂) concepts.

C. *The Cognitive Roles of Acquaintance.* Various meanings of "acquaintance" and of "knowledge by acquaintance" were sorted out in section IV A. Our present concern is with the roles of acquaintance and of knowledge by acquaintance in the enterprise of science, especially in psychology. The first question I wish to discuss concerns the cognitive "plus," i.e., the alleged advantages of knowledge by acquaintance over knowledge by description. We may ask, for example, what does the seeing man know that the congenitally blind man could not know. Or, to take two examples from Eddington (93, 94), What could a man know about the effects of jokes if he had no sense of humor? Could a Martian, entirely without sentiments of compassion or piety, know about what is going on during a commemoration of the armistice? For the sake of the argument, we assume complete physical (1 or 2) predictability and explainability of the behavior of humans equipped with vision, a sense of humor, and sentiments of piety. The Martian could then predict all responses, including the linguistic utterances of the earthlings in the situations which involve their visual perceptions, their laughter about jokes, or their (solemn) behavior at the commemoration. But *ex hypothesi*, the Martian would be lacking completely in the sort of *imagery* and *empathy* which depends on familiarity (direct acquaintance) with the kinds of *qualia* to be imaged or empathized.

As we have pointed out before, "knowledge of," i.e., "acquaintance with," *qualia* is not a necessary condition for "knowledge about" (or knowledge by inference of) those *qualia*. A psychiatrist may know a great deal about extreme states of manic euphoria or of abject melancholic depression, without ever having experienced anything anywhere near them himself. In this case, of course, it must be admitted that the psychiatrist can get an "idea" of these extreme conditions by imaginative extrapolation from the milder spells of elation or depression which he, along with all human beings, does know by acquaintance. But the case is different for observers who are congenitally deprived of acquaintance with an entire modality of direct experience. This is the case of the congenitally blind or deaf, or that of our fancied Martian who has no emotions or sentiments of any kind. But I think it is also

other physical and chemical characteristics of stars and galaxies, etc. Intricate instruments and ingenious theoretical constructions are indispensable in the case of normal (multimodal) perception as well. The difference between persons equipped with all normal sense organs and the deaf-blind is only one of degree, or of the speed with which they would, respectively, attain knowledge about the world in which they are embedded and of which they are parts.

Similar considerations apply to the advantages held by fully equipped persons in regard to psychological and linguistic or descriptive-semantic knowledge. If I have been trained by normal education to apply phenomenal terms (like "red", "green", "lilac fragrance", "rose fragrance", "sweet", "sour", etc.) to qualia of my own direct experience, then I can predict much more readily the application of these terms by other persons in the presence of certain specifiable visual, olfactory or gustatory stimuli. But predictions of this sort are based upon analogical inference; and they are in principle dispensable, because the discriminatory and verbal behavior of other persons is open to intersubjective test. Moreover, if we had a complete neurophysiological explanation of discriminatory and verbal responses we could derive these responses from the cerebral states which initiate them, and which, in turn were engendered by sensory stimulation. Analogously, whatever reliability empathetic understanding in common life, or "clinical intuition" in the psychologist's practice, may have is ultimately to be appraised by intersubjective tests. But the speed with which empathy or intuition do their work depends upon the breadth and the richness of the "experience" of the judge. It also depends upon his use of critical controls.

If the psychologist's personality type is radically different from that of his subject, he will have to correct (often to the point of complete reversal) his first intuitions. For example, an extremely extrovert person will find it difficult to "understand" an extreme introvert, and vice versa. If, however, the personalities are very similar, intuition may "click" readily, and it may even be frequently quite correct. The role of direct acquaintance in all these cases simply amounts to having in one's own experience features and regularities with which one is quite familiar, and which are hence speedily projected and utilized in the interpretation of the behavior of other persons. I conclude that the advantages of direct acquaintance pertain to the context of discovery (cf. Reichenbach, 273) and not to the context of justification. All the examples dis-

cussed do not differ in principle from the obvious examples of persons with "wide experience" as contrasted with persons with "narrow experience," in the most ordinary meaning of these terms. Someone thoroughly familiar with the weather patterns of Minnesota, or with the conduct of business in the Congress of the United States (to take two very different illustrations of the same point) will have the advantages of much speedier inferences and (usually) more reliable predictions than someone who has had no opportunity of long range observations in either case.

The philosophically intriguing questions regarding acquaintance are, I think, of a different sort. They are best expressed by asking, e.g., What is it that the blind man cannot know concerning color qualities? What is it that the (emotionless) Martian could not know about human feelings and sentiments? If we assume complete physical (i.e., at least physical₁) predictability of human behavior, i.e., as much predictability as the best developed physical science of the future could conceivably provide, then it is clear that the blind man or the Martian would lack only acquaintance and knowledge by acquaintance in certain areas of the realm of qualia. Lacking acquaintance means *not having* those experiential qualia; and the consequent lack of knowledge by acquaintance simply amounts to being unable to label the qualia with terms used previously by the subject (or by some other subject) when confronted with their occurrence in direct experience. Now, mere *having* or *living through* ("erleben") is not knowledge in any sense. "Knowledge by acquaintance," however, as we understand it here, is propositional, it does make truth claims; and although it is not infallible, it is under favorable circumstances so reliable that we rarely hesitate to call it "certain." It remains in any case the ultimate confirmation basis of all knowledge claims.

In many of the foregoing discussions we have suggested that what one person has and knows by acquaintance may be identical with what someone else knows by description. The color experiences of the man who can see are known to him by acquaintance, but the blind man can have inferential knowledge, or knowledge by description about those same experiences. After all, this is true as regards an individual color experience even if the other person is endowed with eyesight. The other person does not and could not conceivably have the numerically identical experience (see p. 397f above). Why should we then not conclude that the behavioristic psychologist can "triangulate" the direct experi-

ences of others? I think that indeed he does just that if he relinquishes the narrow peripheralist position, i.e., if he allows himself the introduction of theoretical concepts which are only logically connected with, but never explicitly definable in terms of, concepts pertaining to overt molar behavior. These acquaintancewise possibly unknown states which the behaviorist must introduce for the sake of a theoretical explanation of overt behavior, and to which he (no longer a "radical" behaviorist) refers as the central causes of the peripheral behavior symptoms and manifestations, may well be *identical* with the referents of the phenomenal (acquaintance) terms used by his subject in introspective descriptions of his (the subject's) direct experience. As remarked before, in ordinary communication about our respective mental states, we make this assumption of identity quite unquestioningly. It took a great deal of training in philosophical doubt for learned men to call this assumption into question.

But philosophical doubt, here as elsewhere,* while stimulating in the search for clarity, is ultimately due to conceptual confusions. We have learned how to avoid these confusions, and thus to return with a good philosophical conscience to (at least some of) the convictions of commonsense. We have learned that philosophical doubts, unlike ordinary empirical doubts, cannot be removed by logical or experimental demonstration. What can be demonstrated logically is only the exploitation of certain misleading extensions of, or deviations from, the sensible and fruitful use of terms in ordinary or scientific language. Thus to doubt whether we can at all have knowledge about the "private" experience of other persons is merely the philosophical extension of the ordinary and quite legitimate doubts that we may have in specific instances, for example, when we ask "Is he really as disappointed as his behavior would seem to indicate?" This is to confuse practical difficulties of knowing with (allegedly) basic impossibilities. Once one becomes fully aware of the disease of philosophical skepticism, it becomes possible to cure oneself of it by a sort of self-analysis (logical analysis is what I have in mind here; but in certain cases psychoanalysis may help too, or may even be indispensable).

Granting then that the referents of acquaintance terms and physical₁ theoretical terms may in some cases be identical, this does not by itself

* As, e.g., in the problems of induction, the trustworthiness of memory, the veridicality of perception, etc.

decide the issue between monism and dualism. As we have seen in the previous subsection, the inference to other persons' raw feels can be *logically* differentiated from the inference to their central nervous processes. Dualistic parallelism or epiphenomenalism is entirely compatible with the assertion of the identity of the subjectively labeled mental state with the intersubjectively inferred state which is needed for the explanation of molar behavior. The mental state is logically distinguishable from the "correlated" neurophysiological state. Indeed (as pointed out in section III 4), it makes no sense to talk of *correlation*, or in any case not the usual sense, if the relation of "correlation" were that of *identity*. We shall tackle this crucial point in the next two subsections.

Before we proceed to the discussion of identity and identification, let us however summarize some important conclusions from our discussion of *acquaintance*. The data of direct experience function in three roles: First, in the use of typical patterns and regularities of one person's data for the intuitive or empathetic ascription of similar patterns and regularities of direct experience (or even of unconscious processes) to other persons, these data suggest, but by themselves are never a sufficiently strong basis of *validation* for knowledge claims about the mental life of other persons. Further clinical, experimental, or statistical studies of the behavior of those persons are needed in order to obtain a scientifically respectable degree of confirmation for such inferences. Second, nevertheless, and this is philosophically even more important, the first-person data of direct experience are, in the ultimate epistemological analysis, the *confirmation basis* of all types of factual knowledge claims. This is simply the core of the empiricist thesis over again. But third, the data are also *objects* (targets, referents) of some knowledge claims, viz. of those statements which concern nothing but the occurrence of raw feels or whatever regularities (if any!) can be formulated about raw feels in purely phenomenal terms. For examples of the latter, I mention the three-dimensional ordering of color qualia according to hue, brightness, and saturation; the regularities regarding the gradual (temporal) fading of intense emotions like joy, rage, exultation, embarrassment, regret, grief, etc.; the lawful correlations between, e.g., the experienced contents of daydreams and the attendant emotions of hope or fear. In all these cases, no matter whether the raw feels are our own or someone else's, they are the objects of our knowledge claims or the referents of certain terms in the sentences which describe them. I emphasize this

point because recent empiricist epistemologies in their concern with the confirmation bases of our knowledge claims, and with observation statements which formulate the confirming (or disconfirming) evidence, have tended to neglect consideration of those cases in which the target of the knowledge claim is a state or a regularity of direct experience. Evidence and reference coincide only in the case of statements about the immediate data of first-person experience. But they are clearly distinct in all other cases, such as those in which the object of reference is a state of affairs in the world outside the observer (or else anatomically physiologically *inside* his own skin), no matter whether it be the state of inorganic things, or processes in organisms. Even the direct experience of oneself at a time distinct from the present moment, and of course the direct experience of other organisms or persons are numerically distinct from the data of the confirming evidence. In short, the data of immediate experience function either as *verifiers* or as *referents* of knowledge claims.

D. *Reduction and Identification in Scientific Theories.* In order to decide whether the mental and the physical can in some sense be identified, it is indispensable to cast at least a brief glance at the logic of reduction and identification in the sciences, especially in physics, biology, and psychology. Although these reflections will not provide us with the complete solution of the problem, they will be helpful and suggestive.

It was pointed out and briefly discussed in section II that the advance of scientific theories consists essentially in the reduction of a variety of originally heterogeneous observable facts and regularities to a unitary set of explanatory concepts and postulates. Customarily it is said, for example, that visible light is electromagnetic radiation (within a certain interval of wave lengths); that table salt is NaCl; that magnetized iron is an aggregate of iron atoms with a characteristic spin of certain of their electrons; that the transmitters of hereditary traits are the genes in the chromosomes of the germ cells; that (at least) short range memory traces are reverberating circuits in cerebral cell assemblies, etc. The "is" and the "are" in these sentences represent identities. But these identities differ in their mode of certification from the analytic identities of pure logic and mathematics. For extremely simple illustrations consider the general theorem of set theory " $[S \vee T] = -[-S \cdot -T]$ " or the specific arithmetical identity " $\sqrt{64} = 2^3$ " which hold by virtue of presupposed definitions and the principles of logic or arithmetic.

But the identities established in the factual sciences are confirmed on the basis of empirical evidence. This is very like the empirically ascertainable identity of Shakespeare (or could it be Marlowe?) with the author of *Hamlet*, or the identity of the author of *Hamlet* with the author of *King Lear*. Of course there are also such empirically ascertainable identities as those of Tully and Cicero, of William Thompson and Lord Kelvin, or of the evening star and the morning star. In the examples just given we have (extensional) identities of individuals labeled or uniquely described in two or more ways. When it comes to properties (universals), the identity may be either intensional or extensional. An illustration of the first is, e.g., the identity of $d\sharp$ and $e\flat$ in the well tempered scale of music. An illustration of the second is the identity of the chemical element with atomic number or nuclear charge 20 with calcium characterized as a constituent of limestone, of atomic weight 40, having a melting point of 810°C ., a specific heat of 0.169 at 20°C ., etc.

In the case of analytic identities of individuals or of properties we may speak of the synonymy of names or predicates, respectively. (This applies, of course, also to two-place, three-place, etc. predicates, i.e., to dyadic, triadic, etc. relations. Thus, e.g., "earlier than" is logically synonymous with "temporally precedent to" or with the converse of the relation "later than"). The identity of the class of rational animals with the class of featherless bipeds (disregarding plucked birds), or with the class of laughing animals (disregarding hyenas), is *extensional* and *empirical*. Of course, extensional identity, be it logically necessary or empirical, is implied by intensional identity, but not vice versa. There is no longer any reason to be puzzled about identity being a *relation*. The proper explication of identity consists simply in the recognition that one and the same individual (or universal) may be designated by different labels or described by different characterizations. This could (but need not) be formulated by saying that the relation of identity fully explicated, amounts to a triadic relation between labels (L), or descriptions (D) and a referent (R). The following diagrams represent the simplest paradigmatic situations.

$$L_1 \longrightarrow R \longleftarrow L_2 \text{ or } L \longrightarrow R \longleftarrow D \text{ or } D_1 \longrightarrow R \longleftarrow D_2$$

Since I am not a nominalist, having remained unconvinced by the arguments of Quine, Goodman, and White (269, 242), I see no objec-

tion to introducing *universals* as referents of predicates or relations. And since I am not a Platonic realist either, I am quite willing to consider talk about universals as a convenient *façon de parler*, rather than as a matter of profound "ontological" significance. In my previous example I regarded "d♯" and "e♭" as different labels for the same kind of musical tone-as-heard. Similarly I see no reason whatever to deny that "calcium" and "element of atomic number 20" designate the same kind of substance. This amounts to saying that the identity of universals, if it is not based on the *logical* synonymy of intensions, can amount only to an *extensional* (in this case, empirical) *equivalence* of two classes.

Prima facie the identifications achieved by scientific laws and theories appear to be cases of co-extensiveness, i.e., of extensional equivalence. This is certainly the case with identifications based on empirical laws. A metal characterized in terms of its thermal conductivity may be identical with the metal characterized by its electric conductivity. The ascertainment of the identity, in this case, depends upon the validity of the Wiedemann-Franz law according to which there is a linear relationship between the two kinds of conductivity. Now, while I grant that the word "*identity*" has only one meaning, and this is the meaning defined by the (properly understood) Leibniz principle of *identitas indiscernibilium*, the *modes of ascertainment* of identity are for our purposes the essential consideration. I shall therefore take the terminological liberty of speaking of different kinds of identity, viz., (1) logical, (2) empirical; and under (2) I shall distinguish (a) accidental, (b) nomological, (c) theoretical identities. In more precise but also more cumbersome language this would amount to distinguishing the various modes of ascertainment of identity, or the types of validity that assertions of identity may have.

The identity of the class of rational animals with the class of featherless bipeds may be considered not only as logically contingent, but as empirically accidental; in the same sense as we consider it empirically accidental that the city which is the seat of the United States Government is identical with the city in which on January 17, 1956, at 11:00 a.m. the temperature was (say) 43° F., the barometric pressure 30 inches, and the relative humidity 89 per cent. The referent of these descriptions is the one city of Washington, D.C. This is identity of individuals. Nomological identities rest on empirical laws; theoretical identities depend upon the postulates and definitions of a scientific

theory. Since all types of identity, except the logical, are established on the basis of empirical evidence, they must therefore be formulated in *synthetic* statements.

There is, however, the temptation to regard certain well established *theoretical* identities as analytic. For example, if "gas pressure" is defined in terms of the sum of the momenta delivered by the molecules of a gas to the walls of its container, then of course within the context of the kinetic theory of gases, the identification of pressure with the sum of the molecular momenta is *analytic*. But, as Ernest Nagel (230) has made clear, if we mean by "the pressure of a gas" that property of it which is measurable by manometers, and which has a variety of well-known lawful connections with the volume, the temperature, etc. of the gas, and thus "manifests" itself in a variety of ways, then clearly it was a *discovery*, yielding new information, that revealed to us the relation of gas pressure (the "macro"-concept) to certain aspects of molecular motion. This is clearly *synthetic*. The interesting point which makes it so tempting to view the relation as analytic is, however, worth a little discussion. It is not simply the much vaunted arbitrariness of definitions.* It is rather that the macro-properties and macro-regularities of gases can be *derived*† from the assumptions of the molecular-kinetic theory. A full fledged micro-theory of thermal conduction, convection, diffusion, etc. thus enables us, among other things, to derive the regularities of such indicating instruments as the manometers, thermometers, etc. The expansion of the volume of the gas in the gas thermometer is an immediate logical consequence of the (assumed) increase in the average velocities of the molecules making up the gas, and the initial and boundary conditions which characterize the micro-state of the instrument. Quite analogous considerations apply to the electron theory of electric currents and the measurements of electromotive force and current intensity with the help of such indicating instruments as the voltmeter and the ammeter.

* What is arbitrary in definitions is usually very uninteresting and inconsequential, in contrast to what is not arbitrary.

† It was customary to assume that these derivations are *deductive*. But some of the premises in this case are *statistical* laws; hence some of the derivations of descriptive-observational or empirical-regularity conclusions are probabilistic. Strict deductions, however, can be found in *classical* thermodynamics, *classical* electrodynamics, in the theory of relativity and other examples of "classical" scientific theories. Even in statistical mechanics some derivations are strictly deductive, others so *highly* probable that for practical purposes they can be considered as ("nearly") deductive.

The explanation of the macro-behavior of organisms is sought along methodologically similar lines. Neurophysiological laws and neural-endocrine-muscular, etc. states will presumably suffice for the explanation of even as complex and intricate behavior as that of human beings. Disregarding the ultimately (possibly inevitable) statistical aspects of some of the laws or of the assumptions about initial and boundary conditions, the neurophysiology of the future (3000 A.D.?) should provide complete deductive derivations of the behavior symptoms of various central states whose ψ -correlates are the familiar sensations, perceptions, thoughts, beliefs, desires, volitions, emotions, and sentiments (known by acquaintance and described in phenomenal language). Perhaps I should make clear that I am here trying not so much to convince my readers of the feasibility of what he may consider an entirely utopian and quixotic program for science. I am rather concerned to argue *conditionally*, i.e., if this physicalistic program can be carried out, then there would be something like an *empirical* identification of the referents of molar behavior theory concepts with the referents of some neurophysiological concepts. In its logical and methodological aspects this would be quite analogous to the identification of, e.g., the property of magnetism (as conceived in the macro-theories of physics) with certain micro-structures and processes involving electron spins, etc., ascribed to the atom and quantum dynamics of ferromagnetic substances. These identifications, like all others of a similar kind * appear as analytic only because of the mentioned relations of deducibility which we know (or believe) to hold between the micro-theoretical and macro-nomological or macro-descriptive propositions.

But a more accurate analysis reveals invariably a synthetic-empirical feature somewhere in the context of such scientific explanations. Just where this feature is located depends largely on the nature of the logical reconstruction by means of which we analyze those explanations. In the case of the length of the mercury column in a thermometer, or the volume of the gas in a gas thermometer, the derivation of their (respective) expansions under the condition of increasing heat intensity is so direct that the "identity" appears deceptively as a *logical* one. But even here, empirical regularities enter in. In addition to considerations of the

* E.g., table salt = NaCl; Units of heredity = Genes; Light = electromagnetic waves; the chemical bond = electromagnetic forces playing between the atoms within a molecule; memory traces = reverberating neural circuits; etc., etc.

respective thermal expansion coefficients of gases or mercury as compared with those of the glass of the instruments, there are the laws of geometrical optics regarding the paths of the light rays, and the laws of psychophysics and of psychophysiology concerning the visual perception of the mercury column or of the indicator (e.g., a drop of ink) of the gas thermometer.

Just where we decide to put the boundary (or "partition") between the data of observation and the inferred state of affairs is thus a matter of convenience in epistemological reconstruction. But somewhere we must put it, if we are not to lose sight of the *empirical* character of the relation between the *data* and the *illata*. In one reconstruction the data statements concern the observables of common life. This is the epistemology favored by thinkers like Popper, Carnap, Reichenbach, Hempel, Ryle, Black, Skinner, and W. Sellars. They all agree in this respect even if they differ sharply in others. They all accept in one way or another an intersubjective (physicalistic) thing-language as the basis of epistemological reconstruction. Bertrand Russell, in his later works, is about the only thinker who has made a valiant attempt to combine acceptance of a *phenomenal* basis with a *realistic* (non-phenomenalistic) reconstruction. This means that, as a realist, he has long ago abandoned the earlier phenomenalistic translatability doctrine, and has ever since regarded the relation between physical object statements and phenomenal data statements as one of probabilistic inference. I believe this position still needs considerable logical clarification, but I also believe that it is basically sound, in that it pursues the epistemological analysis down to data which involve only that minimum of inference which knowledge by acquaintance requires. (This was more fully discussed in the preceding subsection.)

No matter where the line is drawn between observables and inferred entities, the most adequate reconstruction, it seems to me, has to be rendered in any case in terms of nomological nets. To return to the temperature example, we may say that the intensity of heat in an oven is indicated by various observable effects, but is not identical with any single one of them, nor is it identifiable with a disjunction (or other logical function) of the observable indications. The intensity of heat is nomologically, and hence *synthetically*, related to the indications of indicators. This is not to be confused with the quite obviously synthetic character of the functional or statistical relations between the indica-

tions themselves. Empiricists, positivists, and operationists have of course always stressed the empirical character of these correlations.

But even when theories (spelling out nomological networks) are adumbrated only in the form of extremely vague "promissory notes," the practice of scientific thinking clearly demonstrates that theoretical concepts (hypothetical entities) are never reducible to, or identifiable with, observable data (or logical constructions thereof). When, e.g., the spirochaete *treponema pallida*, was still undiscovered, the "disease entity" *general paresis* was conceived as the causative factor which "produces" the various symptoms of that disease. Examples of this sort could be multiplied indefinitely from all the sciences. Theoretical concepts are "anchored" in the observables, but are not logically (explicitly) definable in terms of the observables. To be sure, it is the "congruence," "consilience," "convergence," or whatever one wishes to call the testable correlations between the observables that allows for the introduction of fruitful theoretical concepts. It is indeed this consilience which provides the empirical basis for the specification of the meaning of theoretical concepts. Abstract postulates alone determine only their logical or mathematical structure, but never their *empirical* significance.

New evidential bases, such as the microscopic bacteriological findings, provide additional, and usually crucially important, "fixes" upon the theoretical concepts. Nevertheless they amount essentially to enrichments of the nomological net, and thus to a revision of the "weights" of the various other indicators. Thus, in present day pathology, the presence of the spirochaete is a *criterion* of *general paresis*, and even if many of the usual symptoms were absent, the disease would be ascribed to a patient if a sufficient concentration of the spirochaetes in the nerve tissues were verified. The fact that the bacteriological evidence is correlated with the (more "superficial") symptoms is of course something that only observations could have confirmed. But this need not prevent us from saying that the disease entity *general paresis* as construed *before*, or *independently of*, the evidence for the presence of the spirochaete, can be rightfully *identified* with the disease characterized with the help of the bacteriological evidence.

I conclude that it is proper to speak of "*identification*," not only in the purely formal sciences where identity consists in the *logical* synonymy of two or more expressions, but also in those cases in which the mode of ascertainment is empirical. The important consequence for our prob-

lem is then this: Concepts of molar behavior theory like habit strength, expectancy, drive, instinct, memory trace, repression, superego, etc., may yet be identified in a future psychophysiology with specific types of neural-structure-and-process-patterns. The identification, involving as it will, factual discoveries, is *empirical* in its mode of certification, but it is an identification nonetheless.

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 methodologically on a par with the concepts of molar behavior theories (as, e.g., instinct, habit strength, expectancy, drive, etc.) and hence offer in principle no greater difficulties for neurophysiological identification than the concepts of molar behavior

theory which refer to conscious events or processes (e.g., directly experienced sensations, thoughts, feelings, emotions, etc.). As we have repeatedly pointed out, the crux of the mind-body problem consists in the interpretation of the relation between raw feels and the neural processes. The questions to be discussed are therefore these:

1. What does the identity thesis assert about the relation of raw feels to neural events?
2. What is the difference, if there is a difference, between psychophysiological parallelism (or epiphenomenalism) and the identity thesis?
3. Can the identity thesis be defended against empirical arguments which support an interactionistic dualism?
4. Can the identity thesis be defended against philosophical arguments which support dualism on the grounds of the alleged fundamental differences between the properties of direct experience and the features of physical (neurophysiological) processes?

Since I have already paved the way for at least partial replies to question 3, and to some extent also to 4, I shall now primarily concentrate on questions 1 and 2, and discuss the other issues more briefly whenever they will be relevant.

The identity thesis which I wish to clarify and to defend asserts that the states of direct experience which conscious human beings "live through," and those which we confidently ascribe to some of the higher animals, are identical with certain (presumably configurational) aspects of the neural processes in those organisms. To put the same idea in the terminology explained previously, we may say, what is *had-in-experience*, and (in the case of human beings) *knowable by acquaintance*, is identical with the object of *knowledge by description* provided first by molar behavior theory and this is in turn identical with what the science of neurophysiology *describes* (or, rather, will describe when sufficient progress has been achieved) as processes in the central nervous system, perhaps especially in the cerebral cortex. In its basic core this is the "double knowledge" theory held by many modern monistic critical realists.*

* Especially Alois Riehl, Moritz Schlick, Richard Gätschenberger, H. Reichenbach, Günther Jacoby, Bertrand Russell, Roy W. Sellars, Durant Drake, and C. A. Strong. To be sure, there are very significant differences among these thinkers. Russell has never quite freed himself from the neutral monism (phenomenalism) of his earlier neorealistic phase. R. W. Sellars and, following him on a higher level of logical sophistication, his son, Wilfrid, have combined their realistic, double-knowledge view with a doctrine of evolutionary emergence. Opposing the emergence view, Strong and Drake, originally influenced by F. Paulsen, adopted a panpsychistic metaphysics. My own view is a development in more modern terms of the epistemological outlook common

This view does not have the disadvantages of the Spinozistic doctrine of the unknown or unknowable *third* of which the mental and the physical are aspects. The "mental" states or events (in the sense of raw feels) are the referents (denotata) of both the phenomenal terms of the language of introspection, as well as of certain terms of the neurophysiological language. For this reason I have in previous publications called my view a "double-language theory." But, as I have explained above, this way of phrasing it is possibly misleading in that it suggests a purely analytic (logical) translatability between the statements in the two languages. It may therefore be wiser to speak instead of *twofold access* or *double knowledge*. The identification, I have emphasized, is to be *empirically* justified, and hence there can be no *logical* equivalence between the concepts (or statements) in the two languages.

On superficial reflection one may be tempted to regard the identification of phenomenal data with neurophysiological events as a case of the *theoretically* ascertainable identities of the natural sciences. "Theoretical identity" (explicated in section V D) means the sameness of the referent (universal or particular) of two or more *intersubjective* descriptions. For example, it is the atomic micro-structure of a crystal which is indicated ("described") by the optical refraction index, the dielectric constant, the magnetic permeability coefficient, and in greater detail evidenced by X-ray diffraction patterns. Similarly, the various behavioral indications for habit strength refer to a certain, as yet not fully specified, neurophysiological structure in a brain, which may ultimately be certified by more direct histological evidence. Logical Behaviorism admits only intersubjectively confirmable statements and hence defines mentalistic (phenomenal) terms explicitly on the basis of molar behavioral theoretical concepts. Thus, to ascribe to a person the experience of, e.g., an after-image amounts, within the intersubjective frame of reference, to the ascription of a hypothetical construct (theoretical concept), anchored in observable stimulus and response variables. This

to Riehl, Schlick, Russell, and to some extent of that of the erratic but brilliant Gätschenberger. The French philosopher Raymond Ruyer (289, 290) especially before he turned to a speculative and questionable neovitalism (293) held a similar view. Among psychologists W. Köhler (182, 183), E. G. Boring (40), and D. K. Adams (1), again differing in many important respects, hold similar monistic positions. Personally, I consider sections 22-35 in Schlick (298) as the first genuinely perspicacious, lucid and convincing formulation of the realistic-monistic point of view here defended. It is to be hoped that an English translation of this classic in modern epistemology will eventually become available.

theoretical concept may then later be identified, i.e., come to be regarded as *empirically* co-referential with the more detailed and deductively more powerful neurophysiological concept.

The empirical character of the identification rests upon the extensional equivalences, or extensional implications, which hold between statements about the behavioral and the neurophysiological evidence. In our example this means that all persons to whom we ascribe an after-image, as evidenced by certain stimulus and response conditions, also have cerebral processes of a certain kind, and vice versa. In view of the uncertainties and inaccuracies of our experimental techniques we can at present, of course, assert only a statistical correlation between the two domains of evidence. That is to say, the equivalences or implications are, practically speaking, only probabilistic. But in any case, the correlations as well as the theoretical identification of the referents indicated by various items of evidence are formulated in *intersubjectively* confirmable statements.

The identification of raw feels with neural states, however, crosses what in metaphysical phraseology is sometimes called an "ontological barrier." It connects the "subjective" with the "intersubjective." It *identifies* the referents of subjective terms with the referents of certain objective terms. But in my view of the matter there is here no longer an unbridgeable gulf, and hence no occasion for metaphysical shudders. Taking into account the conclusions of the preceding analyses of "privacy", "acquaintance", "physical", and of "identification", private states known by direct acquaintance and referred to by phenomenal (subjective) terms can be described in a public (at least physical₁) language and may thus be empirically identifiable with the referents of certain neurophysiological terms. Privacy is capable of public (intersubjective) description, and the objects of intersubjective science can be evidenced by data of private experience.

The application of phenomenal terms in statements of knowledge by acquaintance is *direct*, and therefore the verification of such statements (about the present moment of subjective experience) is likewise immediate. Phenomenal terms applied to other persons or organisms are used *indirectly*, and the confirmation of statements containing phenomenal terms (thus used) is *mediated* by rules of inference, utilizing various strands in the nomological net as rules of inference. Judging by the structure of one's own experience, there seems to be no reason

to assume the existence of *absolutely* private mental states; i.e., there are presumably no "captive minds" in our world. This is of course a basic ontological feature of nature as we have come to conceive it. It is an *empirical* feature of a very fundamental kind, similar in its "basic frame" character to the 3 + 1 dimensionality of space-time, or to the causal order of the universe. Such frame principles do not differ in kind, although they differ in degree of generality, from the postulates of scientific theories. Their adoption is essentially regulated by the rules of the hypothetico-deductive method.

Logical empiricism as it has come to be formulated in recent years (Carnap, 70, 73; Feigl, 116) recognizes the difference between direct observation (knowledge-by-acquaintance) statements and inferential statements as a contextual difference between direct and indirect confirmation. It does not matter precisely where, in our epistemological reconstruction, we draw the line between the observable and the inferred entities. But wherever we do draw it, the scope of the directly experienceable or of the directly observable depends on the identity of the experiencing and/or observing subject.* What is directly verifiable for one subject is only indirectly confirmable for another. And these very statements (expressed in the preceding two sentences) may be formalized in a pragmatic, intersubjective metalanguage.

Having formulated and in outline explicated the identity thesis, we now have to attend to several important points of philosophical interpretation. I reject the (Spinozistic) double aspect theory because it involves the assumption of an unknown, if not unknowable, neutral ("third") substance or reality-in-itself of which the mental (sentience) and the physical (appearance, properties, structure, etc.) are complementary aspects. If the neutral third is conceived as unknown, then it can be excluded by the principle of parsimony which is an essential ingredient of the normal hypothetico-deductive method of theory construction. If it is defined as *in principle* unknowable, then it must be repudiated as factually meaningless on even the most liberally inter-

* As I understand Dewey and other pragmatists, as well as contextualists like S. C. Pepper (254, 255), this point has been explicitly recognized by them. Cf. also the discussions by analytic philosophers, such as Hampshire (141), Watling (341), and Ayer (18). An exact logical account of the linguistic reflection of direct versus indirect verifiability has been given in the analysis of egocentric particulars (token-reflexive, indexical terms) by B. Russell (286), Reichenbach (274), Burks (58), W. Sellars (308, 312), and Bar-Hillel (20).

puted empiricist criterion of significance. But our view does not in the least suggest the need for a neutral third of any sort. This will now be shown more explicitly.

If a brain physiologist were equipped with the knowledge and devices that may be available a thousand years hence, and could investigate my brain processes and describe them in full detail, then he could formulate his findings in neurophysiological language, and might even be able to produce a complete microphysical account in terms of atomic and subatomic concepts. In our logical analysis of the meanings of the word "physical" we have argued that the physical sciences consist of knowledge-claims-by-description. That is to say that the objects (targets, referents) of such knowledge claims are "triangulated" on the basis of various areas of observational (sensory) evidence. What these objects are acquaintancewise is left completely open as long as we remain within the frame of physical concept formation and theory construction. But, since in point of empirical fact, I am directly acquainted with the qualia of my own immediate experience, I happen to know (by acquaintance) what the neurophysiologist refers to when he talks about certain configurational aspects of my cerebral processes.

There is a danger at this point to lapse into the fallacies of the well-known doctrine of structuralism, according to which physical knowledge concerns only the *form* or *structure* of the events of the universe, whereas acquaintance concerns the *contents* or *qualia* of existence.* This doctrine is to be repudiated on two counts. First, by failing to distinguish acquaintance (the mere *having* of data, or the capacity for imaging some of them) from knowledge by acquaintance (propositions, e.g., about similarities or dissimilarities, rank-orders, etc., of the qualia of the given), the doctrine fails to recognize that even introspective or phenomenological knowledge claims are structural in the very same sense in which all knowledge is structural, i.e., that it consists in the formulation of relations of one sort or another. Second, the realistic interpretation of physical knowledge which we have defended implies that whatever we "triangulate" from various bases of sensory observation is to be considered as "qualitative" in a generalized sense of this term. In the vast majority of cases the qualitative content of the referents of physical descriptions is not "given," i.e., it is not part of a phenomenal field.

* This doctrine has been espoused in various forms by Poincaré (257), Eddington (93), C. I. Lewis (195), Schlick (299), et al.

But it is a given content in the case of certain specifiable neurophysiological processes.

If one wishes to trace the historical origins of this view, one might find it, if not in Aristotle, then certainly in Kant who came very close to saying that the experienced content is the *Ding-an-sich* which corresponds to the brain process as known in the spatio-temporal-causal concepts of natural science.* To put it more picturesquely, in the physical account of the universe as provided in the four-dimensional Minkowski diagram, there are sporadically some very small regions (representing the brains of living and awake organisms) which are "illuminated by the inner light" of direct experience or sentience. This view differs from panpsychism which assumes that the "internal illumination" pervades all of physical reality. But the panpsychists' hypothesis is inconsistent with the very principles of analogy which they claim to use as guides for their reasoning. If one really follows the analogies, then it stands to reason that the enormous differences in behavior (and neural processes) that exist between, e.g., human beings and insects, indicate equally great differences in their corresponding direct experience or sentience. Fancying the qualities of sentience of the lower animals is best left to poetic writers like Fechner, Bergson, or Maeterlinck. As regards the mental life of robots, or of Scriven's (304) "androids," I cannot believe that they could display all (or even most) of the characteristics of human behavior unless they were made of the proteins that constitute the nervous systems—and in that case they would present no puzzle.

The identity view here proposed has met with a great deal of resistance, especially on the part of modern analytic philosophers. To be sure, there are identifications which are "above suspicion." For example, it has been suggested that a legitimate form of empirical identification is to be found in such paradigms as the identity of the "visual" with the "tactual" penny (or the visual, tactual, and olfactory rose; or the visual, tactual, and auditory bell). In each of these examples one may distinguish the various domains of sensory evidence from the particular thing (or thing-kind) that the evidence indicates or refers to. Phenomenalists will, of course, be quick to point out that there is no sense in talking of a thing existing over and above the actual and possible "evidential" data and their important correlations. But from my realistic

* Cf. I. Kant, *Critique of Pure Reason*, section on "The Paralogisms of Pure Reason."

point of view it makes perfectly good sense to explain in terms of physical, psychophysical, and psychophysiological theories how, e.g. a bell by reflecting light, producing sound waves and being a solid, hard body affects our retina, cochlea, and our tactile nerve endings (under specifiable perceptual conditions) and thus produces the visual, tactual, and auditory data in our direct experience. This is indeed the "causal theory of perception" so much maligned by phenomenologists.

We grant that as empiricists we must ultimately justify the causal theory of perception (which is indeed a *scientific* theory, and not an epistemological analysis) by reference to the evidential data which confirm it. And this we can do, no matter whether our own perceptions are concerned (in the egocentric perspective) or those of others (in the "side view" or lateral perspective that we obtain by observing the stimuli, central processes and responses pertaining to other persons). The various sensory "aspects" of the bell are thus to be conceived as the effects which the bell, considered either on the common sense level, or on the micro-level of scientific analysis, has upon our sense organs and finally on our awareness (this last effect empirically identifiable with processes in various cortical areas). Since the phenomenalist thesis of the translatability of physical object statements into data statements is untenable, epistemological analysis must "dovetail" with the causal (scientific) theory of perception and render justice to the latter by an explicit reconstruction of the *nomological* (not purely logical!) relations between the data and the illata. This is still *conceptual* analysis, in that it retraces the relations between the concepts of stimulus objects and the concepts pertaining to the central (cortical-mental) processes in the perceiving organisms.

Our ψ - Φ identification, however, cannot be conceived according to the paradigm of the identity of stimulus objects (like the bell, or the rose). The analogy is misleading in that we have, in the case of stimulus objects physical descriptions of them which together with the empirical laws of psychophysics and psychophysiology enable us (in principle) to derive their various sensory "appearances." Far from requiring an unknown or unknowable "third" or "neutral propertyless substance," ordinary knowledge and especially scientific theory contains a great deal of information about the nature and structure of stimulus objects. The situation in the ψ - Φ case is fundamentally different: We don't have two kinds of evidence for one and the same entity (event, process, etc.). In direct

acquaintance we have, we experience the datum (it is not evidenced, it is evident!), and we identify it with a physical process which we posit as an illatum whose existence is asserted on the basis of multifarious data in other evidential domains.

It should now be clear how the view here proposed differs from the Spinozistic double aspect doctrine. The data of experience are the reality which a very narrow class of neurophysiological concepts denotes. I admit this sounds very "metaphysical." And I shall no doubt be accused of illegitimately extending the ordinary meaning of "denotation". I am fully aware that I am extending the meaning. But I plead that this does not involve my view in paradoxes or needless perplexities. It is true that in common parlance, as well as in the widely accepted philosophical usage, we would say that a term like "neural process in the occipital lobe" denotes a pattern of nerve currents, and not a visual experience. But this remark obviously comes down to the true but trivial semantical assertion that a term designates its designatum; (e.g., "neuron" designates neuron!).

A specification of meaning can be attained through semantic designation rules only if the meaning of the translation equivalent of the definiendum is already understood in the metalanguage. Obviously, according to the commonly accepted usage of the word, a "denotatum" is the referent of proper names, and (except for the null cases) also of predicates, relations, etc. A genuine specification of meaning for empirical terms can be achieved only by a combination of semantical, syntactical and pragmatic rules. The last two types of rules are particularly important. The syntactical rules specify the relations of concepts to one another, and the pragmatic ones make clear which concepts pertain to a basis of direct evidence. The realistic interpretation of empirical concepts depends on an appropriate analysis especially of the roles of proper names (and in scientific languages of coordinates) and of individual-variables (coordinate-variables).*

Taking these analyses into account, we can recognize the valid elements in the older critical realistic epistemology of perceptual and conceptual reference. A physical object or process as perceived in common life, or as conceived in science, is the referent of certain symbolic representations. I submit that it is the preoccupation with the confirmatory

* Cf. especially W. Sellars (308); H. Feigl (110, 111); Bar-Hillel (20).

evidence which has misled positivists and some pragmatists (all of them phenomenologists, radical empiricists, or operationists) to identify the meaning of physical object statements with the actual and/or possible data which, according to our view, merely constitute their evidential bases. Worse still, even sophisticated analytic philosophers tend to confuse the meaning of physical concepts with the perceived or imaged appearance of physical things. No wonder then that we are told that the identity of certain neurophysiological states (or features thereof) with raw feels is a logical blunder. If the denotatum of "brain process (of a specified sort)" is thus confused with the appearance of the gray mass of the brain as one perceives it when looking into an opened skull, then it is indeed logically impossible to identify this appearance with the raw feels, e.g., of greenness or of anxiety.

It would be a similarly bad logical blunder to identify such raw feels with the scientific (heuristic or didactic) tinkertoy models of complex molecular structures (as of amino acids, or proteins) displayed by chemistry instructors in their courses. I don't know whether I should call these blunders "category mistakes." The first one simply consists in the confusion of evidence with the evidenced, or of the indicator with the indicated. What mistake does one make if one confuses smoke with fire, footprints with a man walking, certain darkish spots on an X-ray photograph with tuberculosis? It is strange that of all people it should be the analytic philosophers (who would expose these fallacies with ruthless irony) who do not see that they are making the same sort of mistake in thinking that physical-object concepts denote the perceptual appearance of physical things.

As I have been at pains to point out (in section IV), the only consistent and philosophically fruitful meaning of "physical" (more precisely, of "physical₁") is that of a conceptual system anchored in sensory observation and designed for increasingly comprehensive and coherent explanations of the intersubjectively confirmable facts of observation. This conceptual system or any part of it is in principle non-intuitive (*unanschaulich* as the Germans call it, i.e., unvisualizable). Hence, an identification of a small subset of its referents with something directly given and knowable by acquaintance is in principle left completely open. In point of fact, the imagery commonly, and sometimes helpfully, employed in the thinking of theoretical physicists, biologists, or neurophysiologists consists primarily of pictorial appeals. These are at best

intellectual crutches, fruitful only heuristically or didactically, and not to be confused with conceptual meanings. The fallacy of "introjection" * which was so vigorously criticized by Avenarius (the empiriocriticist of the last century) consists in the pictorial ascription of raw feels to other organisms. As we have seen, such ascriptions indeed clash with the (equally pictorial) ascriptions of physical-appearance properties to other persons or animals.

In the perceptual awareness of other organisms we are confronted with their behavior, i.e., their responses, facial expressions, tone of voice, gait, posture, linguistic utterances, etc., but never with their raw feels. Raw feels do not and cannot be fitted into the appearance picture. They must therefore be conceived as the subjective counterpart of these appearances. As such they are inferentially attainable but not perceptually accessible. At an earlier point we have already discussed the phenomenology of the alleged intuitive or empathetic apprehension of the mental states of other organisms. Since we must recognize intuitive or empathetic ascriptions as fallible and corrigible, they have to be regarded as *inferential* from the point of view of *logical reconstruction* (i.e., in the context of justification), no matter how immediate, "self-evident," compelling, or convincing they may be *psychologically*.

That "introjection" in this sense leads to absurdities becomes especially clear when we consider the ascription of phenomenal fields, e.g., of visual spatiality to other persons. Unless we are solipsists, there is every good reason in the world to ascribe to others the same sort of "life space" (phenomenal environment) which we find so distinctly within our own experience. But if we think of other persons in terms of their appearance in our own phenomenal environment, then it is impossible to ascribe (pictorially) to them also the particular perspectives that they perceive of their environment (or of parts of their own bodies). The fallacy is just as gross as in the case of expecting to find in the brain of another person looking at a green tree a little picture of that tree. But pictorial thinking is one thing, and conceptual thinking is quite another. For conceptual ascription, however, there is no difficulty. The concepts of neurophysiology are non-intuitive and must not be confused with their logically irrelevant pictorial connotations. These connotations lend, psychologically speaking, a certain "root flavor" to

* The term "introjection" as used by R. Avenarius has nothing to do with the well-known homonymous psychoanalytic concept.

these concepts. But once the pictorial appeals connected with the evidential roots of our physical or neurophysiological concepts are dismissed as irrelevant, they no longer pre-empt those places in the conceptual system of which we may then say that they denote some raw feels.

For these reasons I think that once the proper safeguards are applied, no category mistakes are made if we combine phenomenal and physical terms, as indeed we do quite ingenuously not only in ordinary discourse but also in the language of psychology. There is no reason why we should not say, e.g., "The anticipation of success quickened his pace"; "Morbid and tormenting thoughts caused his loss of appetite"; "Touching the hot stove caused intense pain"; "His repressed hostilities finally produced a gastric ulcer"; etc. Category mistakes do arise from confusions of universals with particulars; or of dispositions with occurrents. The first sort of category mistake certainly consists in a violation of the Russellian rule of types. I am not sure whether the second sort can always be reduced to the first. But the original diagnosis made especially by Carnap in his early (phenomenalistic) work (60) of the mind-body perplexities as Russellian-type confusions is no longer acceptable. Physical concepts are not logical constructions out of phenomenal concepts.

A more serious objection to identification comes from reflections upon Leibniz's principle of the identity of indiscernibles. Since we have not only admitted, but repeatedly emphasized the empirical nature of the ψ - Φ identification, one may well ask how we can speak of identity if its confirmation requires the observation of empirical regularities. The most direct confirmation conceivable would have to be executed with the help of an autocerebroscope. We may fancy a "compleat autocerebroscopist" who while introspectively attending to, e.g., his increasing feelings of anger (or love, hatred, embarrassment, exultation, or to the experience of a tune-as-heard, etc.) would simultaneously be observing a vastly magnified visual "picture" of his own cerebral nerve currents on a projection screen. (This piece of science fiction is conceived in analogy to the fluoroscope with the help of which a person may watch, e.g., his own heart action.) Along the lines of the proposed realistic interpretation he would take the shifting patterns visible on the screen as evidence for his own brain processes. Assuming the empirical core of parallelism or isomorphism, he would find that a "crescendo" in his anger, or in the melody heard, would be corresponded by a "crescendo" in the "correlated" cortical processes. (Similarly for "accelerandos," "ri-

tardandos," etc. Adrian's and McCulloch's experiments seem to have demonstrated a surprisingly simple isomorphism of the shapes of geometrical figures in the visual field with the patterns of raised electric potentials in the occipital lobe of the cortex.) According to the identity thesis the directly experienced qualia and configurations are the realities-in-themselves that are denoted by the neurophysiological descriptions. This identification of the denotata is therefore empirical, and the most direct evidence conceivably attainable would be that of the autocerebroscopically observable regularities.

Any detailed account of the ψ - Φ identities is a matter for the future progress of psychophysiological research. But in the light of the scanty knowledge available even today, it is plausible that only certain types of cerebral processes in some of their (probably configurational) aspects are identical with the experienced and acquaintancewise knowable raw feels. A "psychological physiology" * which frames hypotheses about neural structures and processes on the basis of a knowledge of the characteristics and the regularities in the changes of phenomenal fields must therefore always remain extremely sketchy. Knowledge by acquaintance of phenomenal fields alone cannot possibly yield more than a few strands of the total nomological net of neurophysiological concepts required for the explanation of molar behavior. The identification is therefore restricted to those elements, properties, or relations in the neural processes which (in dualistic parlance) are the "correlates" of the raw feels. In our monistic account this is tantamount to the identity of the denotata directly labeled by phenomenal terms, with the denotata of neural descriptions. These latter denotata are acquaintancewise unknown to the neurophysiologist, except if he uses the autocerebroscope himself.

Now it is clear that neural correlates (to speak for the sake of easier exposition once more dualistically) are denoted by concepts which are much richer in meaning than the corresponding phenomenal concepts. The neurophysiological concepts refer to complicated, highly ramified patterns of neuron discharges, whereas their raw-feel correlates may be simple qualities or relations in a phenomenal field. How can, e.g., a uniform patch of greenness, a single musical tone, a stinging pain be identical with a complex set of neural events? Here again it is essential to distinguish between the scientific and the philosophical components of

* Advocated by W. Köhler (184, 185) and critically discussed by C. C. Pratt (260).

this question. Our psychophysiological ignorance is still too great to permit anything more than bold guesses on the scientific side.

There has been talk of "thresholds" and "fusion"; i.e., it is assumed that raw feels emerge only if the intensities of the neural patterns have reached a certain degree; and that complex neural patterns may be "fused" so that the emerging quality "appears" simple and uniform. This sort of talk, though dangerously apt to mislead, is not entirely illegitimate. Talk of thresholds, limens, and fusion is of course quite customary and proper in psychophysics, but its extension to psychophysiology is precarious. It makes perfectly good sense, and is true, to say that the white and black sectors on a swiftly rotating disk phenomenally fuse and yield a uniformly gray appearance. It makes perfectly good sense also, and is equally true, that the intensity of physical stimuli (like light, sound, pressure on one's skin, concentration of chemical substances in the air, etc.) must surpass a certain lower limiting value, if they are to effect a sensation in any of the various modalities (sight, hearing, touch, smell, etc.).

If these facts have any analogies in the intra-cerebral sphere, it would have to be assumed that one area of the cortex "taps" or "scans" other areas and could thus not come to react unless the input reaches a certain intensity. Likewise, one would have to assume that the effect in the second area reflects only certain gross features of the intricate and multifarious process patterns in the first. These would be the analogues of psychophysical thresholds and fusions. Finally, one may assume that the second area (which corresponds to the sensing of the raw feels) is connected with another area corresponding respectively to awareness or judgment (as in introspection) and finally to a motoric area of the cortex which innervates expressive responses or speech.* May I say again that I don't for a moment insist on the scientific adequacy of this particular model. I am not trying to do armchair neurophysiology. All I am concerned to point out is that models are conceivable which would enable us to remove the obstacles arising from the apparent disparities of phenomenal unity versus physical multiplicity; phenomenal spatialities and physical space; phenomenal time and physical time; phenomenal purposiveness and physical causality; etc. I am now going to outline these considerations very briefly.

* I am indebted to R. Carnap for suggesting (in conversations) this sort of brain model.

W. Köhler (182, 183, 185) and R. Ruyer (290, 292, 293) have convincingly shown that the notorious Cartesian perplexities regarding spatiality can be removed by closer attention to the facts of psychophysiology combined with a logical clarification of the distinction between phenomenal space(s) and physical space. (We have laid the groundwork for this in section III B). The surface of objects "physically" outside my skin naturally appears in my visual space as external to the visual appearance of those parts of my body which I can see. There is histological and physiological evidence for a relatively simple projection of the excitation patterns in the retina of the eye, in the area of the occipital lobe of the cerebral cortex. The projection, in its physical and geometrical aspects, is similar to the sort of projection one gets on the screen of a periscope inside a submarine. Not only parts of the surrounding surface of the sea and of other ships, but also parts of the (surfacing) submarine itself are projected upon the screen. Similarly, when I lie on a couch I find not only the appearances of tables, chairs, walls, and windows within my visual field, but I find these object appearances phenomenally outside that part of my phenomenal body (chest, arms, hands, legs, feet) which is also included in my visual field. These simple reflections show that some of the older philosophical puzzles about the outward projection of visual percepts from my mind or brain into the external world are gratuitous, based on confusions, and resolvable by proper attention to the scientific facts on the one hand and to the meanings of spatial terms and phraseologies on the other.

The resolution of the perplexities regarding phenomenal versus physical time, as well as experienced purposiveness versus physical or physiological causality proceeds quite analogously. In the phenomenally temporal "projection" we locate ends-in-view at some distance in the future, and then go about attaining these ends by action, i.e. by the utilization of means. If, e.g., I decide to attend a lecture, I may have to go through a long chain of acts, such as walking to my garage, starting my car, driving to the auditorium, and getting seated there. My actions are clearly goal directed, but there is no need for the myth about the later events (the goal) influencing my antecedent behavior. My behavior is guided, controlled, or modulated by the goal idea which is contemporaneous with my instrumental acts, or possibly precedes them. What in the phenomenal description appears like a future event in my life career determining my current behavior, becomes in the causal account the effect

of one part of my cerebral processes upon another. Of course in this case, just as in the case of memory (recollection), our thinking is essentially mediated by symbols; and therefore "intentionality" (cf. section IV F) plays an important role here. But the symbolic representation of past events or of future events is effected by processes occurring now; i.e., these representations are causal factors in the determination of current behavior. Just as there is no need for a curious notion of "final" causes (or, in Lecomte de Noüy's phrase, of "telefinality"), there is no need for the assumption of a literal presence of the past in present recollections. Whatever the adequate and detailed neurophysiological account of memory traces may ultimately turn out to be, it is these memory traces and not some direct and mysterious apprehension of past events which will causally account for the facts of recollection and of the modification of behavior through learning processes.

Similar considerations would seem to apply to the perennial puzzles concerned with the problems of the nature of the "self," i.e., the unity of the ego, or the unity of consciousness. Here, as in the other puzzles just discussed, the phenomenological descriptions may be correlated with the neurophysiological explanations. Phenomenally there may or may not be a "central core," the "I," in all my experiences. We may admit, following Hume and the later empiricists in the Humean tradition, that there is no distinct element, datum, or impression that could properly be regarded as the self. But it is hard to deny that in the directly given data and in their succession throughout experienced time, there is a certain feature of centralization, coordination, organization, or integration—the reader may choose whichever term seems most suitable. This unitary organization seems to rest on the ever-present potentialities of recollecting a great many events or sequences of events of one's (sic!) past; the ever present possibility of the occurrence of somatic data (referring to one's own body); the existence of a set of dispositions or behavior tendencies, including those ascribed (psychoanalytically) to the superego (i.e., in plain language our set of values and ideals as incorporated in one's conscience); and finally that conception of one's self which is largely a result of the realization of one's own character and personality, adequately or often very inadequately derived from interpretations of one's own behavior and one's social role as perceived by oneself or by others in the social context.

Whichever of these aspects are in some sense phenomenally "given"—

and I suggest a good many may well be so given—these aspects very likely "correspond" to (or according to my view, are identical with) certain relatively stable patterns of cerebral structures and functions. In the pathological cases of split or of alternating personalities (of the Sally Beauchamp, or of the Dr. Jekyll and Mr. Hyde varieties), it has often been suggested that we deal with cerebral subsystems, each having "organic unity" in itself, but only one of them dominating in the determination of behavior during certain intervals of time. If according to psychoanalytic theory large parts of the id as well as of the super-ego are unconscious, this may well be interpreted by assuming that certain portions of the cerebral processes are blocked off (this corresponds to "repressed") from the areas of awareness and of verbal report.

Having rendered plausible the scientific feasibility of at least a parallelistic account of some of the striking and remarkable features of mental life, I return now to the philosophical or logical crux of the identity thesis. We have stressed that the (empirical!) identification of the mental with the physical consists in regarding what is labeled in knowledge by acquaintance as a quale of direct experience as identical with the denotatum of some neurophysiological concept. The scientific evidence for parallelism or isomorphism is then interpreted as the empirical basis for the identification. The step from parallelism to the identity view is essentially a matter of philosophical interpretation. The principle of parsimony as it is employed in the sciences contributes only one reason in favor of monism. If isomorphism is admitted, the dualistic (parallelistic) position may be retained, but no good grounds can be adduced for such a duplication of realities, or even of "aspects" of reality. The principle of parsimony or of inductive (or hypothetico-deductive) simplicity does oppose the operationistic predilection for speaking of two (or more) concepts if the evidential facts, though completely correlated, are qualitatively heterogeneous.

Our view of "triangulation" under such conditions of convergence has, I trust, shown the operationist view to be by far too restrictive. But there is still the logical question how concepts with such fundamentally different evidential bases can be interpreted as (empirically) identifiable. In the case of the concept of the electric current (cf. above section V C) as measured by its magnetic, chemical or thermal, etc. effects, the identification of the several operationally introduced concepts is plausible enough. But, it will again be asked, how can we speak

of identity in the entirely different psychophysiological case where one of the concepts is characterized by the direct applicability of subjective acquaintance terms and the other (the physiological) is introduced on an intersubjective basis and thus has its evidential roots in the sensory data of any qualified observer? I think the answer is not so difficult any more. If we first consider "acquaintance" in its ordinary usage, we can certainly say that Anthony Eden is acquainted with Queen Elizabeth II, and I am not (never having had the opportunity of meeting her). Nevertheless, I can lay claim to some knowledge about the Queen, based on newspaper reports, pictures, and the like. It is surely the same person that Eden and I know, each in his way. Closer to the point, I know by acquaintance what it is to have an eidetic musical-image experience (I occasionally "hear with my inner ear" entire passages from symphonies, string quartets, etc. in their full tone colors). Someone else lacking this sort of experience does not know it by acquaintance, but he can know about it, especially if he is a skillful experimental psychologist. It would be unparisimonious to assume that the psychologist and I are referring to two different (but correlated) processes.

Now, direct acquaintance with "private" raw feels is describable also in the intersubjective language of science. Its ultimate explanation may again have to refer to various cerebral areas, one of which (speaking for ease of exposition again dualistically) "corresponds" to sensing, another to judging, and possibly another yet corresponds to (introspective) reporting. I conclude that acquaintance statements differ only in the type and domain of evidence, but not in regard to their reference, from certain neurophysiological statements. Since the neural apparatus of introspection differs most markedly from that of (external) perception, it should not be surprising that knowledge by acquaintance (now taken in its narrow epistemological sense) is so much more crude, undetailed, and imprecise, than knowledge based on sense perception, especially when this is aided by the instruments of science.

Direct awareness, as we have pointed out before, usually furnishes only qualitative or topological orderings of the contents of phenomenal fields. It could not by itself inform us about the cerebral localization of subjective experience. A very crude (but, if taken literally, I fear highly misleading) analogy might help illuminate this point. A man lost in a jungle perceives the trees and undergrowth in his immediate environment. But the location of this very same part of the jungle can be

determined in a much more accurate and encompassing manner by a cartographer making his measurements from the vantage point of an airplane or balloon high above the jungle. This simile is misleading, of course, in that both the lost wanderer and the cartographer use sensory perception as evidential bases for their knowledge claims. This clearly differs from the case in which I report (or "avow" as Ryle puts it), e.g., a feeling of anxiety and a behavioral psychologist infers my anxiety from the "symptoms," or a neurophysiologist recognizes it in the "corresponding" cerebral processes. Nevertheless, I fail to see that the difference, important though it is in many ways, affects the argument for the identification of the referents of the introspective avowal, with those of the two scientific descriptions.

I conclude that ψ - Φ identity as I conceive it is then still an identity of indiscernibles as defined by Leibniz and Russell. But as the clarification of the "paradox of analysis" (cf. Feyerabend, 120) and of related puzzles about belief sentences should by now have made amply clear, mutual substitutivity even of logically synonymous expressions holds only in non-pragmatic contexts. The empirical synonymy of ψ and Φ terms (or, more cautiously perhaps, their empirical co-reference) a fortiori does not allow for substitutivity in pragmatic contexts. By this I mean that the "salva veritate" condition is fulfilled only in contexts of substitution which do not depend on what we know, or what evidence we have for our knowledge claims. As we pointed out before, there are or were many people (primitive, ancient, etc.) who have no idea of the association of mental life with cerebral processes. But it is nevertheless as justifiable to speak of identity here as it is in the case of "Walter Scott = the author of the Waverley novels," regardless of whether this fact is known or unknown to a given person. In this particular and well-worn example the identity concerns an individual. But, not being a nominalist, I see no difficulties in the identity of a universal, named or described in various ways. Psychophysiological identity may be identity of particulars (this twinge of pain with a specific cerebral event at a certain time), or of universals (pain of a certain kind, and a type of cerebral process).

I am finally going to tackle more specifically and pointedly the question: What is the difference that makes a difference between the parallelism and the identity doctrines? The pragmatist-positivist flavor of this question suggests that it concerns empirically testable differences. But I have already admitted that there are no such differences and

that there could not be any, as far as conceivable empirical evidence is concerned. Is the identity thesis then a piece of otiose metaphysics? Whether it is metaphysics depends of course on what one means by "metaphysics". As I see it, the question is not only similar, but indeed intimately related, to such "metaphysical" issues as realism versus phenomenalism, or the modality versus the regularity view of causality. As most philosophers nowadays realize, these issues, unlike disputes regarding scientific theories cannot be decided by empirical tests. These questions concern the explication of the meaning of concepts and assumptions. They are a subject matter for logical analysis.

As to whether there is a tenable meaning of "causal necessity" related to regularity, but not reducible to it, is a highly controversial issue today. My own reflections favor a view of causal modalities (possibility, necessity, impossibility) which explicates the use of these terms metalinguistically, and nevertheless does not conflict with Hume's basic, and in my opinion irrefutable, contention; viz., that (if I may put it in my own way) the only evidence we can ever have for the assertion of causal connections must be observed regularities. There is, as I see it, no test for causal necessity over and above the tests for regularity. But this does not preclude *meaning* from the distinction between *accidental* and *necessary* universal synthetic statements. A world is conceivable in which a certain metal with a high melting point (say, e.g., platinum) everywhere and always in the infinite history of that world occurs in the solid state, simply because the temperature in that world "happens" never anywhere to surpass a certain upper limit. In such a world the universal statement " $(x,y,z;t)(Pt_{xyz} \supset S_{xyz})$," i.e., "platinum is everywhere and always solid" would be a true universal statement. But the counterfactual conditional "if the temperature were ever to reach or surpass a certain value, platinum would melt" might even be deducible from the basic laws of physics of that world. The universal statement in question is accidentally true. It is not a consequence of a basic law of nature; its truth depends on certain contingent features of the initial and boundary conditions of the fancied world. This shows that there are meaningful distinctions for which no conceivable empirical test could be designed.

Even closer to our problem is the issue between realism and phenomenalism. As I have shown elsewhere (110), there is again no testable difference between these two interpretations of factual knowledge, but there are excellent reasons for the repudiation of phenomenalism and

hence for the acceptance of a realistic epistemology. To relegate the issue to the limbo of metaphysics is a lazy man's way of saving himself the troubles of careful analysis. But close attention to the logic of evidence and reference shows that phenomenalism, even in its most liberal forms does not and cannot substantiate its translatability doctrine; and that only a view which relates phenomenal evidence *synthetically* to statements about physical objects is ultimately tenable.

It is precisely because realists locate both the evidence and the evidenced within the nomological net, that they can give a more adequate account of the relation between "the knower and the known" than positivists, pragmatists, or operationists have ever been able to provide. And it is for this very same reason, that our view of the nature of physical concepts enables us to identify some (of course very few only!) of their referents with the referents of raw feel terms. Dazzled by the admittedly tremendous importance of the evidential basis for our knowledge claims, positivists have regrettably neglected the very *objects* of those knowledge claims. They have myopically flattened them into the surface of evidence, and thus prevented themselves from giving a viable account of the concepts of physics; and they have merely evaded or repressed the mind-body problem which they thought would vanish if their "reductions"—phenomenalistic or behavioristic—were accepted. Ingenious and tempting though their more sophisticated endeavors of reduction have been, they did not succeed. This is why I felt that an explicit reinstatement and defense of a realistic solution of the mind-body problem would be timely and worthwhile.

VI. A Budget of Unsolved Problems. Suggestions for Further Analyses and Research

Although I have proposed what I believe to be at least a fairly circumspect sketch of an adequate solution of the mind-body problems, there are a number of specific component issues which require a great deal of further clarification and investigation. Since I am more interested in the continuing endeavors in this field than in having said the "last word" about it (that's almost inconceivable, in philosophy at any rate!), I shall now attempt to state and discuss succinctly a number of questions to which I have no entirely satisfactory answer at present. I should be immensely pleased if others were to take up these questions in their own work.

The foregoing analyses and discussions were intended to bring to a level of full awareness many of the repressed difficulties of our problem. I have been especially concerned to separate, as well as I could, the scientific from the philosophical issues. And I have tried to show that there are no insuperable logical difficulties for an identity theory of the mental and the physical. I shall again divide the discussion into two parts. The first (A, B, C) will be concerned with open philosophical questions and difficulties. The second (D) will appraise much more briefly the acceptability of identity theory in the light of possibly forthcoming heterodoxical scientific discoveries.

A. *Is There a Phenomenal Language? The Relations of Meaning, Evidence, and Reference.* The central core of the proposed solution rests upon the distinction between evidence and reference. No matter what indirect (behavioral) evidence we use for the ascription of mental states, the mental state ascribed is not to be confused with the evidence which only lends support to the ascription. A fortiori, we must eliminate the still worse confusion of the pictorial appeals (attached to evidential terms) with the conceptual meaning or the reference of neurophysiological concepts. The only case in which pictorial appeals or imagery may be thought to play an essential role in knowledge claims is at the ultimate phenomenal basis of the confirmation of all knowledge claims. And, as we have pointed out, if and only if these knowledge claims are so extremely restricted as to refer exclusively to a currently experienced datum, then—in this very special case—evidence and reference coincide. “Now green”, “now anger”, “now green spot on a gray background”, “stinging pain suddenly increasing”, etc. might be examples. The last example shows that the indexical term “now” need not appear in the phenomenal sentence; but of course the sentence is in the present tense, and this is presumably equivalent with the occurrence of the indexical “now”.

It is difficult to decide whether indexical terms (i.e., egocentric particulars like “now”, “I”, “here”, “this”) are indispensable constituents of singular phenomenal sentences. There are, of course, many examples of universal statements which contain only phenomenal terms as descriptive signs (in addition to purely logical signs): “Orange is more similar to red than it is to green”; “Whatever is colored is extended (in the visual field)”; “Anger always subsides after some time”; etc. There is also the difficult question whether phenomenal sentences

can contain proper names (or something like topological coordinates) for elements in the phenomenal fields. One of my examples suggested that one might use proper names for the small bright spots on the dark background of a visual field and thus describe their relative positions in terms of such relations as “to the left of”, “above”, and “far below”. It seems clear that there is a danger of logical paradoxes, engendered by category mistakes, if we try to mix phenomenal sentences of this sort with the usual behaviorally based ascriptions of mental states to organisms. In these behavioral ascriptions the organism (or the person?) is the individual which is represented by the subject term of the sentence; the predicate is then something like “sees green”, “sees an array of bright spots on a dark background”. There can then be no direct translation of sentences in which the subject terms denote elements in a phenomenal field, into sentences in which the subject terms denote individual organisms. But perhaps there can be an empirical coreference between statements about some (configurational) aspects of neural fields and those about phenomenal fields.

The precise logical explication of empirical identity or coreference is fraught with many difficulties. Some of these stem from the tendency to think of *meaning* as *intension*, and then to conceive of intension in terms of its simplest picturable examples. *Blueness* is an intension indeed, but what are the intensions of “energy”, “entropy”, “electric field strength”, “electric charge”, “neuron discharge”, “reverberating neural circuit”? In all these other cases the intensions are non-intuitive and can be specified only by postulates and correspondence rules. Similarly non-intuitive are the elements of the corresponding extensions, or the denotata. It does seem to me that we can rightly say that both the intension and the extension of the theoretical concepts of the physical sciences are largely unknown by acquaintance, and that only a very small selection of them can therefore be identified with the intensions and extensions of concepts-by-acquaintance. But of course the latter presuppose the existence of a phenomenal language. It has indeed been seriously questioned as to whether there is a phenomenal language at all. In the usual, and full-fledged sense, “language” means a symbolic system with specifiable syntactical (formation and transformation) rules, semantical (designation) rules, and pragmatic (verification) rules. Scraps and bits of phenomenal phraseology seem to fulfill these requirements, but an overall system like that of the physical language does not seem attainable.

The difficulties are further complicated by the question on which level of analysis we are to specify elements and relations described by phenomenal sentences. There is a long history of objections against the Hume-Mach-Russell-Price analysis of experience into "hard" and "soft" data. Phenomenologists, Gestalt psychologists, and more recently many analytic philosophers have raised serious objections not only against the atomism or elementarism of the sense-data doctrine, but also against any doctrine of immediacy or of the given.*

I have throughout this essay maintained and argued that genuinely phenomenal or acquaintance terms are indispensable, not only for the reconstruction of the indirect confirmation of practically all our knowledge claims, but also as labels for the referents of some knowledge claims—whether they are about my own raw feels or that of other humans or animals. I have allowed for the possibility that the "hard data" (i.e., those data which we can talk about with a minimum of inference) are not preanalytically but only postanalytically "given." But on just what level of psychological, introspective, phenomenological, or logical analysis we find those data which stand in the required one-one correspondence to neural events, is an open question. With W. Köhler I am inclined to think that an analysis which stops at a relatively simple configurational level (but does not proceed further to "atomize" the given) may well yield the desired items on the ψ -side of the ψ - Φ isomorphism. But phenomenal description, even of the Gestalt type, is no easy matter.

B. Unitary or Dual Language Reconstruction? In most of the crucial parts of the present essay I have taken a unitary language to be the ideal medium of epistemological reconstruction. By this I mean the following: Both the phenomenal terms (designating raw feel data) and the illata terms (designating unobservables) occur in the language of commonsense or of science, and they are connected by strands in the nomological net. I believe that if this sort of unitary language is constructed with care, category mistakes can be avoided. This reconstruction differs essentially from the dual language reconstruction pursued by Carnap

* For some impressive arguments against atomism see W. Köhler (183, 184, 185), Brunswik (56), Wallraff (340); and against immediacy, Lean (193), Chisholm (75), Wittgenstein (357), Rhees (278), Quinton (270), W. Sellars (315). Others like Ryle (294), Black (38) and Quine (268) have denied the possibility of a phenomenal language altogether. W. Sellars admits phenomenal concepts only as theoretical terms in a language of behavior theory.

and W. Sellars (cf. their essays 73, 315). Purely phenomenal terms are there excluded, presumably owing to their conviction that category mistakes as well as solipsism would be unavoidable if we chose a phenomenal basis of reconstruction. But with the reinstatement of realism, i.e., with the insistence on the synthetic character of the strands in the nomological net, solipsism is no longer a consequence, and category mistakes can be avoided if we dismiss pictorial appeals as cognitively irrelevant, and if we take care to distinguish sharply between universals and particulars, among phenomenal as well as among non-phenomenal terms.

I admit, of course, that there are certain distinct advantages in the dual language reconstruction. All evidential statements are there couched in terms of the observation language; and the observation language is conceived as intersubjectively meaningful right from the beginning. The connections between the observation language and the theoretical language are formulated with the help of correspondence rules. This type of reconstruction is very illuminating in the analysis of the meaning and the confirmation of scientific theories. But, as I have pointed out, it does not do full justice to statements about the data of direct experience, whether they are one's own or someone else's. In our unitary language the "partition" between the data and the illata is located very differently. The correspondence rules in the unitary language would ultimately be statements of ψ - Φ correlations, i.e., of the raw-feel denotations of neurophysiological terms. Since precise knowledge of these correlations is only a matter of hope for a future psychophysiology, the unitary language is largely in the "promissory note" stage. It is therefore not very illuminating if our epistemological reconstruction is to reflect the progress of knowledge in our very unfinished and ongoing scientific enterprise. For this purpose, the dual language reconstruction is much more adequate.

But if we are satisfied with relatively low probabilities for the strands in the nomological net, the unitary reconstruction might do the job too. As a sketch for a reconstruction of an ideally finished science, however, the unitary language approach is preferable. What this would amount to can at present be indicated only by some sort of "science-fiction" illustration: Suppose that we had a complete knowledge of neurophysiology and that we could order all possible human brain states (if not metrically, then at least topologically) in a phase space of n dimen-

sions. Every point in this phase space would then represent a fully specific type of brain state. And, taking isomorphism for granted, a subset of these points would also represent the total set of possible mental states.

Suppose further that we could teach children the vocabulary of the language of brain states. If this requires n -tuples of numbers, then simple expressions like "17-9-6-53-12" (or even abbreviatory symbols for these) might be inculcated in the child's language. If we took care that these expressions take the place of all introspective labels for mental states, the child would immediately learn to speak about his own mental states in the language of neurophysiology. Of course, the child would not know this at first, because it would use the expression, e.g., "17-9-6-53-12" as we would "tense-impatient-apprehensive-yet hopefully-expectant." But having acquired this vocabulary, the child, when growing up and becoming a scientist, would later have no trouble in making this terminology coherent with, and part of, the conceptual system of neurophysiology, and ultimately perhaps with that of theoretical physics. Of course, I not only admit, but I would stress, that in this transformation there is a considerable change in the meaning of the original terms. But this change may be regarded essentially as an enormous enrichment, rather than as a radical shift or a "crossing of ontological barriers." In other words, introspection may be regarded as an approach to neurophysiological knowledge, although by itself it yields only extremely crude and sketchy information about cerebral processes. This sort of information may concern certain Gestalt patterns, certain qualitative and semiquantitative distinctions and gradations; but it would not, by itself, contain any indication of the cerebral connections, let alone localizations.

C. *One-one Correspondence and the "Riddle of the Universe."* The isomorphism of the mental and the physical consists, according to our interpretation, in a one-one correspondence of elements and relations among the phenomenal data with the elements and relations among the referents of certain neurophysiological terms. And we proposed to explain this isomorphism in the simplest way possible by the assumption of the identity of phenomenal data with the referents of (some) neurophysiological terms. The question arises whether the identity view could be held if we were, for empirical reasons, forced to abandon ψ - Φ -one-one correspondence and to replace it by a doctrine of one-many

correspondence. As was pointed out previously, the physicalistic predictability of the occurrence of mental states would in principle still be unique, if one-many correspondence holds true. Comparison with an example of the identification of purely physical concepts may shed some light on this issue. Macro-temperature, as thermometrically ascertained, corresponds in one-many fashion to a multitude of micro-conditions, viz., a very large set of molecular states. Strictly speaking, this correspondence holds between one state description on the macro-level with a specifiable infinite disjunction of state descriptions pertaining to the micro-level. Since, as we have also pointed out, this correspondence is empirically ascertained, there is here as little reason to speak of logical identity as in the ψ - Φ case. Nevertheless, we have seen that it makes sense, and what sense it makes, to regard the relation of temperature to mean molecular kinetic energy as an example of a theoretical identity.

In the mind-body case, just as in the temperature case, prediction of the ensuing micro- (and ultimately even macro-) constellations on the basis of information about, respectively, the mental state, or the macro-temperature state, could not be unique under the supposition of one-many correspondence. This is obvious for the temperature example in the light of the principles of statistical mechanics. Analogously, the precise behavior subsequent to the occurrence of a specified mental state would not be predictable either. This is not too disturbing by itself. After all, even if one-one correspondence held true, the neural correlates of a mental state would form only a very insignificant part of the relevant total initial conditions. Talk of identity in the case of one-many correspondence, however, would seem unjustified, because here we are (*ex hypothesi*) acquainted with the phenomenal datum, and the corresponding disjunction of cerebral states could not plausibly be identified with that individual datum.

Even if one-one correspondence is assumed, there is an intriguing objection* against the identity view. According to the view presented in section V, there is no empirically testable difference between the identity and the parallelism doctrines. We said that the step toward the identity view is a matter of philosophical interpretation. But, so the objection maintains, if identity is assumed, it would be logically impossible to have a stream of direct experience (a "disembodied mind")

* Raised in Minnesota Center for Philosophy of Science discussions by Mr. H. Gavin Alexander.

survive bodily death and decay. It is further asserted that this would not be a logically entailed consequence of parallelism. For it could well be maintained that the one-one correspondence holds only during the life of the person, but that as drastic an event as bodily death marks the limits of this correspondence. Mental states could then occur independently of physical correlates.

Thus it would seem as if our philosophical identity theory implied consequences which are testably different from those of parallelism. This is quite paradoxical. My tentative reply to this argument is two-fold. First, ψ - Φ identification being empirical, it could of course be mistaken. But if the identity does hold, then survival is indeed logically impossible. This is logically quite analogous to the conditional: If the law of the conservation of energy holds, then a *perpetuum mobile* (of the "first kind") is thereby logically excluded. But, of course, the energy law has only empirical validity and might some day be refuted by cogent empirical evidence. Second, and perhaps more important, the parallelism doctrine, as I understand it, holds that there is a ψ - Φ -one-one correspondence and that this correspondence is a matter of universal and irreducible law. This seems to me to exclude disembodied minds just as much as does the identity thesis. I therefore think that the identity thesis is a matter of epistemological and semantic interpretation, and does not differ in empirical consequences from a carefully formulated parallelism.

Another perplexity was formulated in Leibniz's monadology, and in different form presented by E. Dubois-Reymond as one of his famous unsolvable "riddles of the universe." If I may put the core of the puzzle in modern form, it concerns the *irreducible* (synthetic) character of the ψ - Φ correlations. Wherever we find co-existential or correlational regularities in nature, we hope to find a unitary explanation for them, and in many cases scientific theories have provided fruitful and well-confirmed explanations of this sort. But in the case of the ψ - Φ correlation we seem to be confronted with a fundamentally different situation. There is no plausible scientific theory anywhere in sight which would explain just why phenomenal states are associated with brain states. Many philosophers have resigned themselves to regard the ψ - Φ correlations as "ultimate," "irreducible," "brute facts." Since any explanation presupposes explanatory premises which at least in the context of the given explanation must be accepted, and since even the introduction

of higher explanatory levels usually reaches its limit after three or four "steps up," one might as well reconcile oneself to the situation, and say that "the world is what it is, and that's the end of the matter." Now, I think that it is precisely one of the advantages of the identity theory that it removes the duality of two sets of correlated events, and replaces it by the much less puzzling duality of two ways of knowing the same event—one direct, the other indirect.

Nevertheless, there are some "brute facts" also according to the identity theory. But they are located differently. Besides the basic physical laws and initial conditions, there are according to our view the only empirically certifiable identities of denotation of phenomenal and of physical terms. But this identity cannot be formulated in laws or law-like sentences or formulas. The identity amounts merely to the common reference of acquaintance terms on the one hand and unique physical descriptions on the other. Any other way of phrasing the relation creates gratuitous puzzles and avoidable perplexities. For example, it is misleading to ask, "Why does a mental state 'appear' as a brain state to the physiologist?" The brain-state-as-it-appears-to-the-physiologist* is of course analyzable into phenomenal data forming part of the direct experience of the physiologist. The "brute fact" simply consists in this, that the phenomenal qualities known by acquaintance to one person are known (indirectly) by description to another person on the basis of phenomenal (evidential) data which, in the vast majority of cases, are qualitatively quite different from the data had by, or ascribed to, the first person. I see nothing paradoxical or especially puzzling in this account of the matter.

A little reflection upon the autocerebroscopic situation shows clearly that the correspondence between, e.g., musical-tones-as-directly-experienced and certain excitation patterns in the temporal lobes of one's brain as represented by visual patterns (perceived on the screen) is simply a correlation between patterns in two phenomenal fields. The conceptual neurophysiological account of the visual data in this case consists in explanatory hypotheses about cerebral processes which are causally responsible for the production of the image on the screen, and these are in turn causally responsible for the emergence of certain patterns in the visual field. Strictly speaking, and in the light of physical

* No matter whether the physiologist observes someone else's brain, or—autocerebroscopically—his own.

laws, there must even be a minute time lag between the moment of the occurrence of a neural event in the temporal lobe and its "representation" via the autocerebroscope in one's own visual field. The experienced patterns in the visual field are in this situation the causal consequences of (among other things) the auditory data. Disregarding the small time lag we could here speak of a parallelism indeed. But this is a parallelism between the data (or patterns) in different sense modalities; or, in the case of visual experience autocerebroscopically "represented" by other visual data, within one and the same modality. (May I leave it to the reader to think this through and to find out for himself that this special case of autocerebroscopy does not involve any paradoxical consequences.)

Another puzzle that may be raised is the question as to whether the proposed identity theory does not involve the undesirable consequences of epiphenomenalism. It should be obvious by now that our solution of the mind-body problem differs quite fundamentally from materialistic epiphenomenalism in that: (1) it is monistic, whereas epiphenomenalism is a form of dualistic parallelism; (2) the "physical" is interpreted as a conceptual system (or as the realities described by it), but not as the primary kind of existence, to which the mental is appended as a causally inefficacious luxury, or "shadowy" secondary kind of existence; (3) quite to the contrary, mental states experienced and/or knowable by acquaintance are interpreted as the very realities which are also denoted by a (very small) subset of physical concepts. The efficacy of pleasure, pain, emotion, deliberation, volitions, etc. is therefore quite definitely affirmed. In this respect monism shares the tenable and defensible tenets, without admitting the objectionable ones, of interactionism.

Speaking "ontologically" for the moment, the identity theory regards sentience (qualities experienced, and in human beings knowable by acquaintance) and other qualities (unexperienced and knowable only by description) the basic reality. In avoiding the unwarranted panpsychistic generalization, it steers clear of a highly dubious sort of inductive metaphysics. It shares with certain forms of idealistic metaphysics, in a very limited and (I hope) purified way, a conception of reality and combines with it the tenable component of materialism, viz., the conviction that the basic laws of the universe are "physical." This means especially, that the teleology of organic processes, the goal directedness or pur-

posiveness of behavior are macro-features, and that their explanation can be given in terms of non-teleological concepts and laws which hold for the underlying micro-levels. In other words, the monistic theory here proposed does not require irreducibly teleological concepts in its explanatory premises.

In this connection there is, however, a perplexity which may give us pause. Inasmuch as we consider it a matter of empirical fact and hence of logical contingency just which physical (neurophysiological) concepts denote data of direct experience (raw feels), one may wonder whether the causal efficacy of raw feels is satisfactorily accounted for. There are countless teleological processes in organic life which, unless we be panpsychists or psychovitalists, must be regarded as occurring without the benefit of sentience. For examples, consider the extremely "ingenious" processes of reproduction, growth, adaptation, restitution, and regeneration, which occur in lower organisms as well as in many parts of human organisms. On the other hand, the causal efficacy of attention, awareness, vigilance, pleasure, pain, etc. on the human level is so striking that one is tempted, with the panpsychists, to assume some unknown-by-acquaintance qualities quite cognate with those actually experienced.

The new puzzle of epiphenomenalism would seem to come down to this: An evolutionary, physiological, and possibly physical explanation of adaptation, learning, abient or adient, goal-directed behavior can be given without any reference whatever to raw feels. The distribution of raw feels over the various possible neural states could be entirely different from what in fact it is. For example, raw feels might be associated with the peristaltic movements of the stomach or with coronary self-repair, and not with cortical processes. But, I repeat, such different distribution of raw feels or even their complete absence would still not prevent an adequate explanation of teleological behavior. Of course if we accept the actual distribution, i.e., the total set of ψ - Φ -correlation rules as ultimate parallel laws, and interpret these according to the identity theory, then we can quite legitimately speak of the efficacy of raw feels. This is so, because the raw-feel terms are then precisely in those loci of the nomological net where science puts (what dualistic parallelism regards as) their neural correlates. But if the biopsychological explanations offered by the theories of evolution and of learning can thus incorporate the efficacy of raw feels, those theories presuppose, but do not by themselves explain, the ψ - Φ correlations.

That pleasure or satisfaction reinforces certain forms of adient behavior can be formulated in the manner of the law of effect (cf. Meehl, 220). But in the ultimate neurophysiological derivation of this empirical law of behavior, the correlation of pleasure or gratification with certain cerebral states is not required. Behaviorists, especially "logical behaviorists," have taken too easy a way out here in simply defining the pleasurable as the behaviorally attractive and the painful as the behaviorally repellent. The "illumination" of certain physically described processes by raw feels is plainly something a radical behaviorist cannot even begin to discuss. But if the synthetic element in the ψ - Φ relations that we have stressed throughout is admitted, then there is something which purely physical theory does not and cannot account for. Is there then a kind of "brute fact" which our monistic theory has to accept but for which there is possibly no explanation, in the same sense as there can be (within a naturalistic empiricism) no explanation for the fact that our world is what it is in its basic laws and conditions? Possibly, however, I see a riddle here only because I have fallen victim to one of the very confusions which I am eager to eliminate from the mind-body problem. Frankly, I suspect some sort of "regression" rather than "repression" has engendered my bafflement. If so, I should be most grateful for "therapeutic" suggestions which would help in clearing up the issue. Possibly, the solution may be found in a direction which appears plausible at least for the somewhat related puzzle of the "inverted spectrum."

This ancient conundrum, we have seen, is not satisfactorily "dissolved" by Logical Behaviorism. A "captive mind" is logically conceivable, and might know by acquaintance that his sense qualia do not stand in one-one correspondence to his autocerebroscopically ascertained neural states. If physical determinism is assumed, then it is true that such knowledge would have to remain forever private and uncommunicable. But under these conditions a systematic interchange of the qualia for one person at different times and as between different persons is logically conceivable. It would of course *ex hypothesi* not be intersubjectively confirmable, and thus never be a possible knowledge claim of science. But the logical conceivability of the inverted spectrum situation demonstrates again the empirical character of the ψ - Φ correspondence. This empirical character is, however, (as we have also emphasized) extremely fundamental in that it is closely bound up with the basic prin-

ciple of causality or of "sufficient reason." Systematic interchange of qualia for the same sort of neural states would be something for which, *ex hypothesi*, we could not state any good reasons whatever.

Furthermore, there is a grave difficulty involved in the assumption that a captive mind could even "privately" know about the interchange. Normal recollection by memory presumably involves (at least) quasi-deterministic neural processes. The captive mind could be aware of the inverted spectrum type of interchange of qualia only if we assume some peculiar breach in normal causality. If the captive mind is to know that today the correlation of raw feels with neural states differs from what it was yesterday, he would have to remember yesterday's correlations. But how could this be possible if memory depends upon modifications in the neural structures of the cortex? These considerations show clearly that under the supposition of normal physical causality the systematic interchange would remain unknowable even to the private captive mind. (Converse, but otherwise analogous, puzzles arise for the assumption of the survival of a private stream of experience beyond bodily death. How could such a private mind have knowledge about the continuance of his "physical" environment?)

All these reflections seem to me to indicate that in our world at least, there is nothing that is in principle inaccessible by "triangulation" on an intersubjective (sensory) basis. The having of raw feels is not knowledge at all, and knowledge by acquaintance does not furnish any truths which could not in principle also be confirmed indirectly by persons other than the one who verifies them directly. The ψ - Φ -identity theory as I understand it, makes explicit this "ontological" feature of our world. The criterion of scientific meaningfulness formulated in terms of intersubjective confirmability, far from being an arbitrary decree or conventional stipulation, may thus be viewed as having ontological significance—but "ontological" in the harmless sense of reflecting an inductively plausible, basic characteristic of our world.

Empirical identity, as I conceive it, is "weaker" than logical identity but "stronger" than accidental empirical identity, and like theoretical identity stronger than nomological identity in the physical sciences (just as causal necessity is weaker than logical necessity, but stronger than mere empirical regularity). If the coreference of a phenomenal term with a neurophysiological term is conceived as something more than mere extensional equivalence, if it is conceived as characteristic of the

basic nature of our world (just as the basic natural laws characterize our kind of world and differentiate it from other kinds), then perhaps the inference from a neural state to its ("correlated") raw feel is at least as "necessary" (though of course not purely deductive) as is the inference from, e.g., the atomic structure of a chemical compound to its macro-physical and chemical properties.

I hope that readers sympathetic to my admittedly speculative gropings will try to formulate in logically more precise and lucid form what I have been able to adumbrate only so vaguely. Such readers should in any case keep in mind one of the ideas which seem to me indispensable for an adequate solution of the phenomenalism-realism as well as the mind-body problems: The paradigm of symbolic designation and denotation is to be seen in the relation of a token of a phenomenal term to its raw-feel referent. All non-phenomenal descriptive terms of our language, i.e., all physical terms (no matter on which level of the explanatory hierarchy) designate (or denote) entities which—within the frame of physical knowledge—are unknown by acquaintance. But if our "hypercritical" realism is accepted, we must ascribe denotata to all those physical terms which designate individuals, properties, relations, structures, fields, etc., i.e., entities which can justifiably be said to be described (i.e., uniquely characterized) on the basis of evidential data by Russellian descriptions on one or the other level in the hierarchy of logical types. "To exist" means simply to be the object of a true, uniquely descriptive statement. But since such descriptive knowledge (on a sensory evidential basis) by itself never enables us deductively to infer the acquaintance qualities of its objects, there is always a possibility for some sort of modal identification of a datum with a specifiable *descriptum*. This is the central contention of the present essay.

D. *Some Remarks on the Philosophical Relevance of Open Scientific Questions in Psychophysiology*. There are many problems of predominantly scientific character among the various mind-body puzzles. These await for their solution the further developments of biology, neurophysiology, and especially of psychophysiology.* We have touched on

* The following works and articles strike me as especially important, or at least suggestive, in these fields: Boring (40); Köhler (183, 184); Wiener (349); Hebb (145); Herrick (154); Adrian (3, 4, 5); Brain (46); Eccles (92); Ashby (9); McCulloch (214, 215); von Foerster (122, 123); Blum (39); Brillouin (49); Culbertson (80); Colby (76); Gellhorn (132); Krech (188, 189). Northrop's (240) exuberant and enthusiastic appraisal of the significance of cybernetics for the mind-body prob-

many of these issues in various parts and passages of the present essay. Speaking (again for ease of exposition only) the language of parallelism, there are, e.g., the following issues to be decided by further research concerning the specific ψ - Φ correspondences:

1. The problem of the cerebral localization of mental states and functions: Classical and recent experiments indicate quite specific localization for many processes. On the other hand, the findings of Lashley, Köhler, and others demonstrate a principle of mass action or of the equipotentiality of various cerebral domains.

2. The problem of the relation of phenomenal (visual, tactual, kinesthetic, auditory, etc.) spatialities to physical space: The time-honored puzzle regarding (Lotze's) "local signs" is, as far as I know, not completely resolved. The question is by what neural mechanisms are we able to localize narrowly circumscribed events (like sensations of touch or of pain) more or less correctly on our skin or within our organism? Can we assume projection areas in the cortex which through learning processes come to interconnect afferent neural impulses in the different sensory modalities, and thus enable us to localize, e.g., visually what is first given as a tactual or pain sensation?

3. The problem of the nature of memory traces: Current fashion makes much of the reverberating circuits in neural structures. But it seems that while this explanation may do for short-range memory, it is probably not sufficient for long-range memory. Whether the lowering of neural or synaptic resistance is to be explained by "neurobiotaxis," by thickenings of the bud ends of dendrites, or by some chemical (quantum-dynamical) change in the neurons, is at present quite dubious.

4. The problem of the "specious present": The fact that the direct experience of one conscious moment embraces the events in a short stretch of finite duration, and not just an "infinitesimal" of physical time, presents a puzzle that is intriguing especially from a philosophical point of view. It is difficult, but I think not impossible to conceive of scanning mechanisms which "take note" of freshly accumulated traces, and even involve an extrapolative aspect as regards the immediate future.

5. The problem of the recollection of ordered sequences of past experiences: How can a brain process at a given time provide a correct lem indicates at least one philosopher's response to the challenge of this new borderland discipline.

simultaneous representation of such a sequence? Philosophers are used to distinguishing a sequence of remembrances from the remembrance of a sequence of events. It seems that the latter can in certain instances occur in one moment of the specious present. Thus I seem to be aware of the sequence of themes and developments in the first movement of Beethoven's Seventh Symphony, and this awareness does not seem to require a quick internal rehearsal. It seems to be "all there at once." I also can, usually with fair reliability, recall the temporal sequence of many events in my life (various voyages, lecture engagements, first, second, third, etc. visits to Paris, and so on). Is it again some sort of "scanning" mechanism which might account for this? Driesch (87) considered it outright impossible to conceive any neurophysiological mechanism which would explain these phenomena, and believed that only a dualistic interactionism (involving a strictly immaterial mind or self, consonant with the rest of his vitalistic doctrines) could render justice to them. While I know of no obviously workable neural model that would do the trick, I think that Driesch, here as elsewhere, declared the defeat of naturalistic explanations prematurely. Present-day scientific findings and scientific theorizing have in so many cases shown the feasibility of physicochemical explanations of biological phenomena, so that we have good reasons to expect a successful solution of the problem of remembrance of past event-sequences.

6. The problems of "quality," "fusion," and "thresholds": I have dealt with these as best I could above (section V E), but there is no doubt that future research is needed in order to provide an adequate explanation for these striking phenomena.

7. The problems of "wholeness" (Gestalt), teleological functioning and purposive behavior: These also were discussed above (section IV E). The contributions of Gestalt theory and its doctrine of isomorphism have been largely absorbed in current psychophysiology (cf. especially Hebb, 145). Similarly significant and hopeful are the analyses of negative feedback processes as provided by cybernetics. The doctrines of "General Systems Theory," though related in spirit to cybernetics, Gestalt theory, and mathematical biophysics, are however very dubious from a logical point of view (cf. Buck, 57). We have also discussed the related issue of emergent novelty. If "absolute emergence" (Pap, 244) is a fact, then perhaps some such account as that given by Meehl and Sellars (221) may be considered seriously. I still expect that future

scientific research will demonstrate the sufficiency of physical₂ explanations. But if I should be wrong in that, a theory involving genuine emergence would seem to be a much more plausible alternative than dualistic interactionism. Such a theory would, however, have important philosophical implications. Inference to mental states would rest on presupposed nomological relations between physical₂ brain events and mental states which could be defined only in terms of the theoretical concepts of a physical₁ language. There would still be empirical identity between the referents of some (theoretical) physical₁ terms and the referents of phenomenal terms, but the scientific explanation of behavior would be markedly different from purely physical₂ explanations. Some of the philosophical puzzles of the mind-body problem might be resolved even more plausibly under this hypothesis. For example, the question regarding the "inverted spectrum" could be answered, quite straightforwardly, on the basis of normal inductive or analogical inference. Directly given qualia, represented by (theoretical) physical₁ terms in our scientific account would then be functionally related to those brain processes which are described in physical₂ (theoretical) terms. The principle of sufficient reason would then tell us that to assume any deviation from the highly confirmed functional relationships between mental states and physical₂ brain states would be just as arbitrary as, e.g., the assumption that some electric currents are associated with magnetic fields of an entirely different structure than are others (despite the complete similarity of the electric currents in every other respect). As I have indicated before, the validity of the emergentist theory falls in any case under the jurisdiction of future empirical research.

8. The problem of a neurophysiological account of *selfhood*: This important though controversial notion describes a form of organization or integration of experiences and dispositions which on the neural side corresponds first to the relatively stable structure of the brain and the other parts of the nervous system, as well as to certain unified forms of functioning. To what extent the psychoanalytic concepts of the ego, superego, and id may be "identified" with such structures and functions is still very unclear. Very likely, the psychological notions will appear only as first crude approximations, once the detailed neurophysiological facts are better known.

9. The problems of neurophysiological theories which will account for the unconscious processes assumed by various "depth psychologies,"

especially psychoanalysis: One of the philosophically intriguing questions here is whether we can explicate such psychoanalytic concepts as "repressed wishes", "unconscious anxiety", "Oedipus complex", etc. as dispositions, or whether unconscious events also need to be assumed. Even outside the sphere of Freudian preoccupations, there are for instance the often reported cases of "waking up with the solution of a mathematical problem." One wonders whether the brain did some "work" during sleep, and if so, whether "unconscious thoughts" might not be part of a first-level explanation of this sort of phenomenon. I am inclined to think that both dispositions and events are required, and that the future development of science may well produce more reliable neurophysiological explanations than the currently suggested (and suggestive) brain models (cf. Colby, 76).

10. Much more problematic than all the questions so far discussed in this section are the implications of the alleged findings of psychical research. Having been educated in the exercise of the scientific method, I would in the first place insist on further experimental scrutiny of those findings. But if we take seriously the impressive statistical evidence in favor of telepathy, clairvoyance, and precognition, then there arises the extremely difficult problem of how to account for these facts by means of a scientific theory. I know of no attempt that gives even a plausible suggestion for such a theory. All hypotheses that have been proposed so far are so utterly fantastic as to be scientifically fruitless for the present. But logical analyses (e.g., C. D. Broad, 52; M. Scriven, 304) which make explicit in which respects the facts (if they are facts!) of psychical research are incompatible with some of the guiding principles of ("Victorian"!) science are helpful and suggestive. It is difficult to know whether we stand before a scientific revolution more incisive than any other previous revampings of the frame of science, or whether the changes which may have to be made will only amount to minor emendations.

Concluding Remark. An essential part of the justification of the philosophical monism proposed in this essay depends upon empirical, scientific assumptions. Only the future development of psychophysiology will decide whether these assumptions are tenable. Since I am not a laboratory scientist (though I did some laboratory work in physics and chemistry in my early years), I cannot responsibly construct psychophysiological hypotheses. Nor did I intend to close the doors to

alternative philosophical views of the relations of the mental to the physical. What I did try to show, however, is that monism is

- (1) still very plausible on scientific grounds,
- (2) philosophically defensible in that it involves no insurmountable logical or epistemological difficulties and paradoxes.

I realize fully that I could deal only with some of the perplexities which have vexed philosophers or psychologists throughout the ages, and especially in recent decades. Just where the philosophical shoe pinches one, just which problems strike one as important—that depends, of course, on a great many more or less accidental personal, educational, or cultural factors. Despite my valiant efforts to deal with what strike me as important and baffling questions, I may of course not even have touched on other facets which some of my readers might consider as the essential problems of mind and body. May others come and deal with them!

NOTE AND REFERENCES

Since this essay has almost the dimensions of a monograph, I feel I should acknowledge my sincere indebtedness to the countless philosophers and scientists who have helped me by their publications as well as (in many instances) by personal discussion or correspondence to reach whatever clarity I may claim to have achieved. It is impossible to mention them all, but some stand out so distinctly and prominently that I should list them. Naturally, I have learned from many of these thinkers by way of disagreement and controversy. In any case none of them is to be held responsible for whatever may be wrong or confused in my views. My first acquaintance with philosophical monism goes back to reading the work of Alois Riehl (279); I found essentially the same position again in Moritz Schlick (298), some of whose work I had studied before I became his student in Vienna in 1922. I have profited enormously (although he may well think, not sufficiently) from discussions with my kind and patient friend R. Carnap intermittently throughout more than thirty years. During my Vienna years (1922–30) I was greatly stimulated by discussions also with Schlick, Wittgenstein, Victor Kraft, Otto Neurath, E. Kaila, Karl Popper, Edgar Zilsel, et al. I was greatly reinforced in my views by my early contact with the outstanding American critical realist C. A. Strong (in Fiesole, Italy, 1927 and 1928). Along similar lines I found corroboration in the work of Roy W. Sellars, Durant Drake, Richard Gätschenberger, and in some of the writings of Bertrand Russell. Discussions (and many controversies) during my American years, beginning in 1930, with E. G. Boring, S. S. Stevens, P. W. Bridgman, C. I. Lewis, A. N. Whitehead, H. M. Sheffer, V. C. Aldrich, S. C. Pepper, E. C. Tolman, C. L. Hull, B. F. Skinner, K. Lewin, E. Brunswik, W. Köhler, Albert Einstein, H. Reichenbach, F. C. S. Northrop, and Philipp Frank, proved most stimulating.

During the last three and a half years of the activities of the Minnesota Center for Philosophy of Science I had the tremendous advantage of intensive discussions not only with my colleagues Paul E. Meehl, Wilfrid Sellars, and Michael Scriven, each of whom disagrees with me on several different fundamental points, and each for different reasons, but I also profited from discussions with such visitors or collaborators as Gilbert Ryle, C. D. Broad, Anthony Flew, Peter Strawson, Ernest Nagel, C. G. Hempel, A. Kaplan, Arthur Pap, Herbert Bohnert, Henry Mehlberg, Hilary Putnam,