

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

The "Mental" and the "Physical"

Herbert Feigl

E. Arguments Concerning the Identification of Sentience with Neural Events

I shall now present, as explicitly as I can, the reasons for an empirical identification of raw feels with neural processes. I shall also discuss several apparent 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.¹

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 theoretical concept may then later be identified, i.e., come to be regarded as empirically correlative 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 physical) 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, 1950, 1956; Feigl, 1963) 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.² 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 interpreted 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 acquaintanceshipwise 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.³ 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. 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.⁴ 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 (1953) "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.

NOTES

1. 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 to Riehl, Schlick, Russell, and to some extent of that of the erratic but brilliant Gätschenberger. The French philosopher Raymond Ruyer (1930, 1934)

especially before he turned to a speculative and questionable neovitalism (1952) held a similar view. Among psychologists W. Köhler (1929, 1938), E. G. Boring (1933), and D. K. Adams (1954), again differing in many important respects, hold similar monistic positions. Personally, I consider sections 22–35 in Schlick (1925) 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.

2. As I understand Dewey and other pragmatists, as well as contextualists like S. C. Pepper (1950, 1960), this point has been explicitly recognized by them. Cf. also the discussions by analytic philosophers, such as Hampshire (1952), Watling (1954), and Ayer (1956). An exact logical account of the linguistic re-

flection of direct versus indirect verifiability has been given in the analysis of egocentric particulars (token-reflexive, indexical terms) by B. Russell (1940), Reichenbach (1947), Burks (1949), W. Sellars (1948, 1954), and Bar-Hillel (1954).

3. This doctrine has been espoused in various forms by Poincaré (1929), Eddington (1928), C. I. Lewis (1929), Schlick (1938), et al.
4. Cf. I. Kant, *Critique of Pure Reason*, section on "The Paralogisms of Pure Reason."

REFERENCES

- Adams, D. K. "Learning and explanation," *Learning Theory, Personality Theory, and Clinical Research: The Kentucky Symposium*, pp. 66–80. New York: Wiley; London: Chapman & Hall, Ltd., 1954.
- Ayer, A. J. *The Problem of Knowledge*. New York: St. Martin's Press, 1956.
- Bar-Hillel, Y. "Indexical Expressions," *Mind*, 63:359–379 (1954).
- Boring, E. G. *The Physical Dimensions of Consciousness*. New York, London: The Century Co., 1933.
- Burks, A. W. "Icon, Index and Symbol," *Philosophical and Phenomenological Research*, 9:673–89 (1949).
- Carnap, R. "Empiricism, Semantics, and Ontology," *Revue Internationale de Philosophie*, 4:20–40 (1950). Reprinted in P. P. Wiener (ed.), *Readings in the Philosophy of Science*, pp. 509–21. New York: Scribner, 1953. Also reprinted in L. Linsky (ed.), *Semantics and the Philosophy of Language*, pp. 208–30. Urbana (Ill.): Univ. of Illinois Press, 1952.
- Carnap, R. "The Methodological Character of Theoretical Concepts," in *Minnesota Studies in the Philosophy of Science*, Vol. I, pp. 38–76. Minneapolis: Univ. of Minnesota Press, 1956.
- Eddington, A. S. *The Nature of the Physical World*. New York: Macmillan, 1928.
- Feigl, H. "Physicalism, Unity of Science and the Foundations of Psychology," in P. A. Schilpp (ed.), *The Philosophy of Rudolf Carnap*, LaSalle (Ill.): Open Court Pub. Co., 1963.
- Hampshire, S. "The Analogy of Feeling," *Mind*, 61:1–12 (1952).
- Köhler, W. "Ein altes Scheinproblem," *Die Naturwissenschaften*, 17:395–401 (1929).
- Köhler, W. *The Place of Values in a World of Facts*. New York: Liveright, 1938.
- Lewis, C. I. *Mind and the World Order*. New York: Scribner, 1929.
- Pepper, S. C. "The Issue Over the Facts," in *Meaning and Interpretation*. Berkeley: Univ. of California Press, 1950.
- Pepper, S. C. "A Neural Identity Theory," in S. Hook (ed.), *Dimensions of Mind*, pp. 37–56. New York: New York Univ. Press, 1960.
- Poincaré, H. *The Foundations of Science*. New York: The Science Press, 1929.
- Reichenbach, H. *Elements of Symbolic Logic*. New York: Macmillan, 1947.
- Russell, B. *An Inquiry into Meaning and Truth*. New York: Norton & Co., 1940.
- Ruyer, R. *Esquisse d'une Philosophie de la Structure*, Paris: F. Alcan, 1930.
- Ruyer, R. "Les Sensations Sont-elles dans Notre Tête?" *Journal de Psychologie*, 31:555–80 (1934).
- Ruyer, R. *Néo-Finalisme*. Paris: Presses Univ. de France, 1952.
- Schlick, M. *Allgemeine Erkenntnislehre*. Berlin: Springer, 1925.
- Schlick, M. *Gesammelte Aufsätze*. Vienna: Gerold & Co., 1938.
- Scriven, M. "The Mechanical Concept of Mind," *Mind*, 62:230–40 (1953).
- Sellars, W. "Realism and the New Way of Words," *Philosophical and Phenomenological Research*, 8:601–34 (1948). Reprinted in H. Feigl and W. Sellars (eds.), *Readings in Philosophical Analysis*, pp. 424–56. New York: Appleton-Century-Crofts, 1949.
- . "Some Reflections on Language Games," *Philosophy of Science*, 21:204–228 (1954).
- Watling, J. "Ayer on Other Minds," *Theoria*, 20:175–80 (1954).