

it seems unlikely that any physical theory of mind can be contemplated until more thought has been given to the general problem of subjective and objective. Otherwise we cannot even pose the mind-body problem without sidestepping it.

## 13

---

### Panpsychism

By panpsychism I mean the view that the basic physical constituents of the universe have mental properties, whether or not they are parts of living organisms. It appears to follow from a few simple premises, each of which is more plausible than its denial, though not perhaps more plausible than the denial of panpsychism.

#### 1. *Material composition*

Any living organism, including a human being, is a complex material system. It consists of a huge number of particles combined in a special way. Each of us is composed of matter that had a largely inanimate history before finding its way onto our plates or those of our parents. It was once probably part of the sun, but matter from another galaxy would do as well. If it were brought to earth, and grass were grown in it, and milk from a cow that ate the grass were drunk by a pregnant woman, then her child's brain would be partly composed of that matter. Anything whatever, if broken down far enough and rearranged, could be incorporated into a living organism. No constituents besides matter are needed.

#### 2. *Nonreductionism*

Ordinary mental states like thought, feeling, emotion, sensation, or desire are not physical properties of the organism – behavioral, physiological, or otherwise – and they are not implied by physical properties alone.<sup>1</sup>

<sup>1</sup> Strictly speaking, the argument requires only that some mental states are not reducible.

3. *Realism*

Nevertheless they are properties of the organism, since there is no soul, and they are not properties of nothing at all.<sup>2</sup>

4. *Nonemergence*

There are no truly emergent properties of complex systems. All properties of a complex system that are not relations between it and something else derive from the properties of its constituents and their effects on each other when so combined. Emergence is an epistemological condition: it means that an observed feature of the system cannot be derived from the properties currently attributed to its constituents. But this is a reason to conclude that either the system has further constituents of which we are not yet aware, or the constituents of which we are aware have further properties that we have not yet discovered.

Panpsychism seems to follow from these four premises. If the mental properties of an organism are not implied by any physical properties but must derive from properties of the organism's constituents, then those constituents must have nonphysical properties from which the appearance of mental properties follows when the combination is of the right kind. Since any matter can compose an organism, all matter must have these properties. And since the same matter can be made into different types of organisms with different types of mental life (of which we have encountered only a tiny sample), it must have properties that imply the appearance of different mental phenomena when the matter is combined in different ways. This would amount to a kind of mental-chemistry.

The conclusion has its attractions as a general explanation of how conscious life arises in the universe. But there are three problems about the argument that I want to discuss.

1. Why call these inferred properties of matter mental? What is meant by a physical property and why does that concept not apply to them?
2. What view of causality is involved in the denial of emergence?

<sup>2</sup> Some of them, like belief and perception, are relational properties, but all involve some nonrelational aspect.

3. Do the features of mental phenomena that argue against reduction also argue against Realism?<sup>3</sup>

To deal with the first question, we must consider what makes a newly discovered property or phenomenon physical. Since the class of known physical properties is constantly expanding, the physical cannot be defined in terms of the concepts of contemporary physics, but must be more general. New properties are counted as physical if they are discovered by explanatory inference from those already in the class. This repeated process starts from a base of familiar, observable spatio-temporal phenomena and proceeds to take in mass, force, kinetic energy, charge, valence, gravitational and electromagnetic fields, quantum states, anti-particles, strangeness, charm, and whatever physics will bring us next.<sup>4</sup>

What the argument claims is that a similar chain of explanatory inference beginning from familiar mental phenomena would lead to general properties of matter that would not be reached along the path of explanatory inference by which physics is extended. Let us put aside for the moment the uneasiness that one may well feel about the suggestion that mental phenomena should derive from any properties of matter at all.

The claim is that if such properties exist, they are not physical in the sense explained. No properties of the organism or its constituents discovered solely by physics will be the familiar mental properties with their conscious or preconscious aspects, nor will they be the more basic proto-mental properties that imply these; for it will never be legitimate to infer, as a theoretical explanation of physical phenomena alone, a property that includes or implies the consciousness of its subject. We do infer explicitly mental explanations of physical behavior, but these employ concepts understood independently and not introduced through physical theory. Theories constructed on the basis of physical observations and parallels alone will not include terms that imply the consciousness of the system.

<sup>3</sup> I shall capitalize this term when using it in the special sense of premise 3.

<sup>4</sup> This is roughly equivalent to Feigl's 'physical<sub>2</sub>'. See H. Feigl, 'The "Mental" and the "Physical"', *Minnesota Studies in the Philosophy of Science*, vol. II, ed. H. Feigl, M. Scriven, and G. Maxwell (Minneapolis: University of Minnesota Press, 1958).

It is this assumption about inference that underlies the position that the physical will never include the mental. If it is true, then in the event that any properties of matter are discoverable by explanatory inference from observable mental phenomena, they will have mental implications of a kind that physically inferred properties will never have. In that sense the ultimate properties inferred to explain mental processes would be mental and not physical.

However, this needs modification, for there is a third possibility. Perhaps there are not two chains of inference, but one chain leading from the mental and the physical to a common source. It is conceivable in the abstract that if mental phenomena derive from the properties of matter at all, those may be identical at some level with nonphysical properties from which physical phenomena also derive.

This merits a brief digression. Such reducibility to a common base would have the advantage of explaining how there could be necessary causal connexions in either direction, between mental and physical phenomena. It would also make less problematic the possibility that a single event like a bodily movement could have both a mental cause and a complete physical explanation. The mental cause, sufficiently analyzed, could be part of the physical cause, sufficiently analyzed. But if this were so, the common reducing properties would not be physical. They could not be reached by a chain of explanatory inference from physical phenomena alone, for physical data alone would provide no grounds for postulating explanatory theories that also had mentalistic consequences. The theories that physical data provide grounds for may take extraordinary leaps which permit the deduction of radical physical consequences (the convertibility of matter and energy, the deflection of light by gravity). But without any mentalistic evidence there is no reason to give mental content to the explanation of physical events. (Someone who infers from a drought that the rain god is angry is not basing his hypothesis on physical evidence alone. He is making a psychological interpretation of the drought, based on familiarity with human motivation. Any inference of this kind, reasonable or unreasonable, does not belong to physics.) Therefore even if there are common ultimate properties underlying both the mental and the physical, they do not lie on the path of physical

discovery, the path of explanatory inference from observable physical phenomena alone, and so they are not physical properties.

If there were such properties, they would be discoverable only by explanatory inference from both mental and physical phenomena. This seems in fact somewhat less implausible than that there are two quite distinct chains of explanation leading back to two distinct sets of basic properties. If it were true, then it would be improper to describe the basic properties as mental for the same reason that they could not be described as physical. Strictly, only what is inferred to explain mental phenomena (including actions) should be called mental. This clearly admits concepts like repression and utility function, or perhaps universal grammar.<sup>5</sup> They appear at a level of psychological theory not far removed from familiar mental processes. But even if by some criterion the fundamental particles had properties that were not mental but neither mental nor physical, the conclusion of the argument would survive in a modified form. There would be properties of matter that were not physical from which the mental properties of organic systems were derived. This could still be called panpsychism.

The second question is about causality and emergence. What is the view of causal explanation from which it follows that true emergence is impossible? I have said that the properties of a complex system must *derive* from the properties of its constituents, plus the way they are combined. The argument assumes that uniform correlations cannot provide an adequate basis for the explanation of complex phenomena. It therefore rejects what is often called, inaccurately, a Humean analysis of causation. According to Hume, our idea of causal necessity is a kind of illusion, because all we ever observe are natural regularities and correlations, and never necessary connexions of cause and effect. Hume did not think that our idea of cause was that of an instance of a constant conjunction in nature.

<sup>5</sup> I have discussed the sense in which such concepts of psychological theory are mental in 'Linguistics and Epistemology', in *Language and Philosophy*, ed. Sidney Hook (New York: New York University Press, 1969), and in 'Freud's Anthropomorphism', in *Freud*, ed. Richard Wollheim (New York: Doubleday, 1974).

He was right, in my opinion, to say that if this were all there was, then causality would be an illusion. But I do not believe it is an illusion. True causes *do* necessitate their effects: they make them happen or make them the case. Uniform correlations are at best evidence of such underlying necessities. This seems to me clearly true in elementary cases: heat causing water to boil, rocks causing glass to break, magnets inducing electric current, the wind making waves. Given what heat is and what water is, it is literally impossible for water to be heated beyond a certain point at normal atmospheric pressure without boiling.

Causal necessity operates even at the most fundamental levels. An electron is a particle with a certain charge and a certain mass. Those properties imply that it will interact in a definite way with fields and with other objects. Some of the implications will be probabilistic, but that does not affect the point. And similar things are true of other subatomic particles. Ordinary physics and chemistry explain macroscopic phenomena, so far as they can be explained, as the necessary consequences of the properties of the particles (sometimes essential properties) and their interactions. They do not rely merely on contingent correlations.

This is particularly clear when we consider the relation between properties of complex systems and properties of their components *at the same time*. Consider the physical properties of a diamond. Some of them, like shape, size, weight, and crystal structure, are directly entailed by the physical properties and relations of its constituents and their effects on each other when they are so combined. Others, like color, glitter, and hardness, involve interaction between the diamond and other things, and must be explained in terms of the effects of the diamond's constituents on those other things.

The supposition that a diamond or an organism should have truly (not just epistemologically) emergent properties is that those properties appear at certain complex levels of organization but are not explainable in terms of any more fundamental properties, known or unknown, of the constituents of the system. If causal connexions were nothing but instances of contingent regularities, such a situation would be compatible with the existence of causal explanations of the emergent properties at a complex level. There would probably be many uniform psycho-physical correlations of the form; 'Whenever an

organism is in *exactly* physical state P it is also in mental state M.' This may be true of my present total physical and mental states, for example. No doubt more general correlations also exist.

On a correlation view that should be enough for M to be causally explained by P. But it is not enough on a stronger view of causation. A stronger view requires that P somehow *necessitate* M; but at this complex level, no necessary connexions can be discovered. There is no sense in which my body's physical state *by itself* makes it the case that I am in mental state M. It is of course obvious that what is going on in my brain causes my mental state, just as it is obvious that when I touch a hot pan it causes pain. There *must* be some kind of necessity here. What we cannot understand is *how* the heat, or the brain process, necessitates the sensation. So long as we remain at the level of a purely physical conception of what goes on in the brain, this will continue to appear impossible. The conclusion is that unless we are prepared to accept the alternative that the appearance of mental properties in complex systems has no causal explanation at all, we must take the current epistemological emergence of the mental as a reason to believe that the constituents have properties of which we are not aware, and which do necessitate these results.

The demand for an account of how mental states necessarily appear in physical organisms cannot be satisfied by the discovery of uniform correlations between mental states and physical brain states, though that is how psycho-physical laws have traditionally been conceived. Instead, intrinsic properties of the components must be discovered from which the mental properties of the system follow necessarily. This may be unattainable, but if mental phenomena have a causal explanation such properties must exist, and they will not be physical.<sup>6</sup>

The third question, about Realism, is the most difficult. What is the reason to deny that mental properties can be entailed by physical ones? It is certainly conceivable that the physiological and behavioral characteristics of a living organism should follow necessarily from the physical properties of fundamental particles

<sup>6</sup> The inference to such properties is not trivial, like the statement that opium puts people to sleep because it has a dormative virtue. Although the causes are formulated so as to entail their effects, the reverse implication does not hold, as it does in the joke.

when they are combined in that way, though we can never expect to possess more than fragments of such an explanation. This is true also of functional states, so called, if they are defined in terms of their relations to one another, to stimuli, and to behavior. If the definition is general enough, the functional state could appear in a wide variety of physical systems, including organisms whose behavior took widely different forms. But its presence could still be entailed by the physical micro-properties of any organism in which it appeared.

A physical explanation of behavioral or functional states does not explain the mental because it does not explain its subjective features: what any conscious mental state is like for its possessor. Let me say briefly what I mean by this, though it is too large a topic for proper discussion here.<sup>7</sup> A feature of experience is subjective if it can in principle be fully understood only from one *type* of point of view: that of a being like the one having the experience, or at least like it in the relevant modality. The phenomenological qualities of our own experiences are subjective in this way. The physical events in our brains are not. Human physiologists may take a special interest in them; but they can, in principle, be understood just as well, or even better, by creatures totally unlike us in physical and mental structure. To understand them such creatures need not take up our point of view. Physical brain processes can be understood objectively, from the outside, because they are not subjective phenomena. And no description or analysis of the objective nervous system, however complete, will ever by itself imply anything which is not objective, i.e. which can be understood only from one kind of viewpoint, that of the being whose states are being described. One cannot derive a *pour soi* from an *en soi*.

Not all mental states are conscious, but all of them are capable of producing states that are. So any derivation of the mental properties of an organism from the properties of its components would have to show that subjective states necessarily arise from them. Of course if, as was suggested earlier, the explanation of behavior leads ultimately to properties that are neither mental nor physical, then a sufficiently basic explanation of the physical aspects of behavior might also explain subjective experience as a

<sup>7</sup> I try to give a fuller account of this idea in chapter 12 above.

necessary part of the process. But physical properties alone could not give this result; they explain not how things are from a particular subjective point of view but how they are objectively, in ways that can be apprehended from different points of view and do not belong to any.

This gap is logically unbridgeable. If a bodiless god wanted to create a conscious being, he could not expect to do it by combining together in organic form a lot of particles with none but physical properties.<sup>8</sup> Given an account of the phenomenology of a particular kind of perception, it may be possible to deduce how a particular objective state of affairs would appear from that point of view. But the subjective premise seems essential. And this is no less true when the objective state is a physical brain state, and the appearance is what it is like to be *in* that brain state, rather than what it is like to observe it.

That, in brief, is the argument against reductionism. Because of the way in which it relies on the subjectivity of the mental, I believe that it casts doubt on Realism, though I find this hard to explain.

For Realism as I have defined it to be true, physical organisms must have subjective properties. What seems unacceptable about this is that the organism does not have a point of view: the person or creature does. It seems absurd to try to discover the basis of the point of view of the person in an atomistic breakdown of the organism, because that object is not a possible subject for the point of view to which the person's experiences appear. And if it makes no sense to ascribe subjective states to the complex whole, there will be no basis for ascribing proto-mental states to its constituents; so *they* cannot be appealed to in explanation of what it means for an organism to have experiences. I simply record this feeling of impossibility because I have no more to say about it. When a mouse is frightened it does not seem to me that a small material object is frightened.

The trouble with this intuition is that it leads nowhere. What is the alternative? I assume that neither I nor the mouse has a soul, to bear these mental properties. And even if we did, it would not remove the problem, because insofar as it is possible

<sup>8</sup> Cf. Saul Kripke, 'Naming and Necessity', in *Semantics of Natural Language*, ed. D. Davidson and G. Harman (Dordrecht: Reidel, 1972), pp. 340-1.

to grasp the idea of a nonmaterial thing, there is just as much difficulty in understanding how it could have a point of view. But if the occurrence of a subjective experience is not the possession of a property by *something*, what is it? And what connexion does it have with the organism? Evidently in some way experiences depend on the material organism even if they are not states of it.

The only view I know of that may qualify as an alternative is found in the *Philosophical Investigations*. According to Wittgenstein as I understand him the person (or mouse) who is the subject of mental states is not to be identified with an organism or a soul or anything else. He holds that all kinds of familiar propositions about the mental states of individual living beings are true, but that there is almost nothing to be said about what property must be possessed by what thing if one of these ascriptions is to be true. All such specifications of truth conditions are trivial. What can be more fully described, however, are the kinds of circumstances, including evidential grounds, that make the ascription appropriate: criteria rather than truth conditions. For third-person ascriptions the grounds are behavior, stimuli, circumstances, and testimony (once the subject has learned the relevant mental vocabulary). For self-ascriptions no evidential grounds are needed.

Although facts about the body are among the criteria for ascribing mental states to others, and also for ascribing to them an understanding of the terms they use to ascribe mental states to themselves, the mental states are not states of the body. The view is not reductionist. Mental states are no less real than behavior, physical stimuli, and physiological processes. In fact their situation with respect to one another is symmetrical, because physical processes have mental (specifically observational) criteria just as mental processes have physical criteria. According to Wittgenstein, everything there is must be systematically connected with other things in a way that permits public agreement, or at least public disagreement, about whether it is there or not. Mental phenomena meet this condition through their connexion with behaviour and circumstances, but they are perfectly real in their own right. They cannot be analyzed as dispositions to behavior or properties of the organism, any more than physical phenomena can be analyzed as

multiple possibilities of sensation or of observation. If asked to say what any of these kinds of thing really *is*, or what statements about them really *assert*, we can give no reply that is not trivial.

In some ways that is an attractive position. It does justice to the subjectivity of the mental, because of the central place it assigns to criterionless mental self-ascriptions. How things appear to someone must hang together with how they appear to others to appear to him; but these facts are inextricably connected with his point of view, as this can be publicly identified. There is clear support for the idea that mental states are subjective if they are ascribed to creatures who can ascribe them to themselves without observation, by other creatures who can ascribe similar states to *themselves* in the same way. And since it does not seem correct to describe these states of the individual as states of the organism, this idea provides an alternative to Realism.

My difficulty with the view is that it depends too heavily on our language. Essentially its account of mental phenomena is an account of how they are ascribed, particularly in the first person. But not all conscious beings are capable of language, and that leaves the difficult problem of how this view accommodates the subjectivity of *their* mental states.

We ascribe experience to animals on the basis of their behavior, structure, and circumstances, but we are not just ascribing to them behavior, structure, and circumstances. So what are we saying? The same kind of thing we say of people when we say they have experiences, of course. But here the special relation between first- and third-person ascription is not available as an indication of the subjectivity of the mental. We are left with concepts that are anchored in their application to humans, and that apply to other creatures by a natural extension from the behavioral and contextual criteria that operate in ordinary human cases.

This seems definitely unsatisfactory, because the experiences of other creatures are certainly independent of the reach of an analogy with the human case. They have their own reality and their own subjectivity. They are not, I assume, of indeterminate character in cases where the natural extension from human behavior and circumstances gives no determinate result. To take a very clear case, if things emerged from a spaceship which we

could not be sure were machines or conscious beings, what we were wondering about would have an answer even if the things were so different from anything we were familiar with that we could never discover it. It would depend on whether there was something it was like to be them, not on whether behavioral similarities warranted our saying so.

This seems true quite apart from the question of what the subject of mental states is. They may not be states of the body, but they certainly exist in forms beyond the reach of our language. So they cannot be analyzed in terms of human criteria for their ascription. And since human experiences have the same kind of reality, must not the same be true of them? What they are is not fully captured by an account of the conditions under which first- and third-person ascriptions of experience are appropriate.

I will mention that this raises problems about whether the concept of experience, as I am applying it, meets basic conditions of publicity that it must meet to be well-defined at all. It is widely accepted that one cannot always define a type of similarity or a type of thing simply by pointing to an instance and saying 'the same as *this*'. And it may be doubted whether someone who wonders whether the things coming out of the spaceship have experience, without any idea of the possibility of determining whether they do or not, is really asking a well-defined question. I think that in this case the conditions of meaning are met, but I will not try to defend the claim here. Experience must have systematic connexions with behavior and circumstances in order for experiential qualities and experiential similarity to be real. But we need not know what these connexions are in order to ask whether experience is present in an alien thing.

I therefore seem to be drawn to a position more 'realistic' than Wittgenstein's. This may be because I am drawn to positions more realistic than Wittgenstein's about everything, not just the mental. I believe that the question about whether the things coming out of the spaceship are conscious must have an answer. Wittgenstein would presumably say that this assumption reflects a groundless confidence that a certain picture unambiguously determines its own application. That is the picture of something going on in their heads (or whatever they have in place of heads)

that cannot be observed by dissection.

Whatever picture I may use to represent the idea, it does seem to me that I know what it means to ask whether there is something it is like to be them, and that the answer to that question is what determines whether they are conscious – not the possibility of extending mental ascriptions on evidence analogous to the human case. Conscious mental states are real states of something, whether they are mine or those of an alien creature. Perhaps Wittgenstein's view can accommodate this intuition, but I do not at the moment see how.

Where does this leave us? I have now expressed dissatisfaction with three alternative interpretations of mental states: that they are states of the body, that they are states of the soul, and that all we can say about their essence is to give criteria or conditions for their ascription. But what is left? If they are real states of something in the world, if they depend on what is going on in the creature's body, if they are intimately connected with stimuli and behavior, and if the creature does not consist of a body plus something else, what can experience be but states of the organism? If one accepts realism in a broad sense about mental states, it is very difficult to avoid Realism in the more specific sense that forms a premise of the argument for panpsychism.

This of course expresses that fatal step in the philosophy of mind, the argument by elimination. There is no reason to think that all possibilities have been thought of, so there is no reason to assume that a view is correct if all currently conceivable alternatives are even more unacceptable. Still, when a mouse or a fly or a man comes to exist because matter has been combined in certain ways, the resulting mental states seem to have to belong to the organism for want of a better home. Realism may be the weakest premise in the argument, but it is more plausible at the moment than its denial.

I therefore believe that panpsychism should be added to the current list of mutually incompatible and hopelessly unacceptable solutions to the mind-body problem. It can be avoided by denying any of the premises of the argument. Denial of the first results in dualism. This still leaves problems about the causal connexions between mind and body: either (a) those connexions are pure correlations and not necessary; or (b) the body will have properties that necessitate mental effects in the soul and effects of

the soul on the body; or else (c) the soul will have properties that enable it to be acted on by the body and vice versa. If (b), then the body will have mental or at least non-physical properties. If (c), then the soul will have physical properties as well as mental ones.

Denial of the second premise is fairly common among contemporary philosophers, but the only motive I can see for accepting any of the resulting kinds of reductionism is a desire to make the mind-body problem go away. None of them has any intrinsic plausibility.

Denial of the third premise, Realism, is more attractive but awaits the development of a viable alternative, some way of admitting the reality of mental occurrences without ascribing them to either organisms or souls as subjects.

Denial of the fourth premise, nonemergence, involves accepting the existence of irreducible contingent laws connecting complex organic states with mental states. In a sense this would mean that mental states had no causal explanation: that they were not necessitated by anything. I do not believe the world is like that, but here, as with the other premises, one *can* take that escape-route. It would be useful to develop all the alternatives more fully.

As for panpsychism, it is difficult to imagine how a chain of explanatory inference could ever get from the mental states of whole animals back to the proto-mental properties of dead matter. It is a kind of breakdown we cannot envision, perhaps it is unintelligible. Presumably the components out of which a point of view is constructed would not themselves have to have points of view. (How could a single self be composed of many selves?) Yet they would have to be recombinable to form different points of view, for not only can a single organism have different experiences, but its matter can be recombined to form other organisms with totally different forms of experience. The mental properties of all matter, therefore, would have to be not species-specific but universal, since they would underlie all possible forms of consciousness. In a sense, they would be less subjective than any of the specific forms.

Panpsychism in this sense does not entail panpsychism in the more familiar sense, according to which trees and flowers, and perhaps even rocks, lakes, and blood cells have consciousness of

a kind. But we know so little about how consciousness arises from matter in our own case and that of the animals in which we can identify it that it would be dogmatic to assume that it does not exist in other complex systems, or even in systems the size of a galaxy, as the result of the same basic properties of matter that are responsible for us.<sup>9</sup>

<sup>9</sup> My ideas on this topic, especially on the concept of the physical and the role of necessity in causal explanation, have been strongly influenced by Rebecca Goldstein and William L. Stanton. Their own views are developed in Stanton's 'Anomalous Monism and The Mental Qua Mental' (Ph.D. dissertation, Princeton University, 1975) and Goldstein's 'Reduction, Realism, and Mind' (Ph.D. dissertation, Princeton University, 1976).