

Naturalistic Dualism

DAVID CHALMERS

Nonreductive Explanation

There are principled reasons to think that no reductive explanation of consciousness, in terms of underlying physical processes, is possible. As I argued in chapter 17, an account of physical processes may solve some of the easy problems of consciousness, concerning the explanation of cognitive functions, but it can never explain the existence of conscious experience. At this point some are tempted to give up, holding that we will never have a theory of conscious experience. McGinn (1989), for example, argues that the problem is too hard for our limited minds; we are “cognitively closed” with respect to the phenomenon. (For further discussion, see Rowlands, chapter 26.) Others have argued that conscious experience lies outside the domain of scientific theory altogether.

I think this pessimism is premature. This is not the place to give up; it is the place where things get interesting. When simple methods of explanation are ruled out, we need to investigate the alternatives. Given that reductive explanation fails, *nonreductive* explanation is the natural choice.

Although a remarkable number of phenomena have turned out to be explicable wholly in terms of entities simpler than themselves, this is not universal. In physics, it occasionally happens that an entity has to be taken as *fundamental*. Fundamental entities are not explained in terms of anything simpler. Instead, one takes them as basic, and gives a theory of how they relate to everything else in the world. For example, in the nineteenth century it turned out that electromagnetic processes could not be explained in terms of the wholly mechanical processes that previous physical theories appealed to, so Maxwell and others introduced electromagnetic charge and electromagnetic forces as new fundamental components of a physical theory. To explain electromagnetism, the ontology of physics had to be expanded. New basic properties and basic laws were needed to give a satisfactory account of the phenomena.

Other features that physical theory takes as fundamental include mass and space-time. No attempt is made to explain these features in terms of anything simpler. But this does not rule out the possibility of a theory of mass or of space-time. There is an intricate theory of how these features interrelate, and of the basic laws they enter into. These basic principles are used to explain many familiar phenomena concerning mass, space, and time at a higher level.

I suggest that a theory of consciousness should take experience as fundamental. We know that a theory of consciousness requires the addition of *something* fundamental

to our ontology, as everything in physical theory is compatible with the absence of consciousness. We might add some entirely new nonphysical feature, from which experience can be derived, but it is hard to see what such a feature would be like. More likely, we will take experience itself as a fundamental feature of the world, alongside mass, charge, and space-time. If we take experience as fundamental, then we can go about the business of constructing a theory of experience.

Where there is a fundamental property, there are fundamental laws. A nonreductive theory of experience will add new principles to the furniture of the basic laws of nature. These basic principles will ultimately carry the explanatory burden in a theory of consciousness. Just as we explain familiar high-level phenomena involving mass in terms of more basic principles involving mass and other entities, we might explain familiar phenomena involving experience in terms of more basic principles involving experience and other entities.

In particular, a nonreductive theory of experience will specify basic principles telling us how experience depends on physical features of the world. These *psychophysical* principles will not interfere with physical laws, as it seems that physical laws already form a closed system. Rather, they will be a supplement to a physical theory. A physical theory gives a theory of physical processes, and a psychophysical theory tells us how those processes give rise to experience. We know that experience depends on physical processes, but we also know that this dependence cannot be derived from physical laws alone. The new basic principles postulated by a nonreductive theory give us the extra ingredient that we need to build an explanatory bridge.

Of course, by taking experience as fundamental, there is a sense in which this approach does not tell us why there is experience in the first place. But this is the same for any fundamental theory. Nothing in physics tells us why there is matter in the first place, but we do not count this against theories of matter. Certain features of the world need to be taken as fundamental by any scientific theory. A theory of matter can still explain all sorts of facts about matter, by showing how they are consequences of the basic laws. The same goes for a theory of experience.

This position qualifies as a variety of dualism, as it postulates basic properties over and above the properties invoked by physics. But it is an innocent version of dualism, entirely compatible with the scientific view of the world. Nothing in this approach contradicts anything in physical theory; we simply need to add further *bridging* principles to explain how experience arises from physical processes. There is nothing particularly spiritual or mystical about this theory – its overall shape is like that of a physical theory, with a few fundamental entities connected by fundamental laws. It expands the ontology slightly, to be sure, but Maxwell did the same thing. Indeed, the overall structure of this position is entirely naturalistic, allowing that ultimately the Universe comes down to a network of basic entities obeying simple laws, and allowing that there may ultimately be a theory of consciousness cast in terms of such laws. If the position is to have a name, a good choice might be *naturalistic dualism*.

If this view is right, then in some ways a theory of consciousness will have more in common with a theory in physics than with a theory in biology. Biological theories involve no principles that are fundamental in this way, so biological theory has a certain complexity and messiness to it; but theories in physics, insofar as they deal with fundamental principles, aspire to simplicity and elegance. The fundamental laws of nature are part of the basic furniture of the world, and physical theories are telling us that this basic furniture is remarkably simple. If a theory of consciousness also involves fundamental principles, then we should expect the same. The principles of simplicity, elegance, and even beauty that drive physicists' search for a fundamental theory will also apply to a theory of consciousness.

(A technical note: Some philosophers argue that even though there is a *conceptual gap*

between physical processes and experience, there need be no metaphysical gap, so that experience might in a certain sense still be physical (Levine 1983 and chapter 29; Loar 1990; Hill 1991). Usually this line of argument is supported by an appeal to the notion of a posteriori necessity (Kripke 1980). I think that this position rests on a misunderstanding of a posteriori necessity, however, or else requires an entirely new sort of necessity that we have no reason to believe in (see Chalmers 1996; also Jackson 1994 and Lewis 1994 for details). In any case, this position still concedes an *explanatory* gap between physical processes and experience. For example, the principles connecting the physical and the experiential will not be derivable from the laws of physics, so such principles must be taken as *explanatorily* fundamental. So even on this sort of view, the explanatory structure of a theory of consciousness will be much as I have described.

Outline of a Theory of Consciousness

It is not too soon to begin work on a theory. We are already in a position to understand certain key facts about the relationship between physical processes and experience, and about the regularities that connect them. Once reductive explanation is set aside, we can lay those facts on the table so that they can play their proper role as the initial pieces in a non-reductive theory of consciousness, and as constraints on the basic laws that constitute an ultimate theory.

There is an obvious problem that plagues the development of a theory of consciousness, and that is the paucity of objective data. Conscious experience is not directly observable in an experimental context, so we cannot generate data about the relationship between physical processes and experience at will. Nevertheless, we all have access to a rich source of data in our own case. Many important regularities between experience and processing can be inferred from considerations about one's own experience. There are also good indirect sources of data from observable cases, as when one relies on the verbal report of a subject as an indication of experience. These methods have their limitations, but we have more than enough data to get a theory off the ground.

Philosophical analysis is also useful in getting value for money out of the data we have. This sort of analysis can yield a number of principles relating consciousness and cognition, thereby strongly constraining the shape of an ultimate theory. The method of thought-experimentation can also yield significant rewards, as we will see. Finally, the fact that we are searching for a *fundamental* theory means that we can appeal to such nonempirical constraints as simplicity, homogeneity, and the like in developing a theory. We must seek to systematize the information we have, to extend it as far as possible by careful analysis, and then make the inference to the simplest possible theory that explains the data while remaining a plausible candidate to be part of the fundamental furniture of the world.

Such theories will always retain an element of speculation that is not present in other scientific theories, because of the impossibility of conclusive intersubjective experimental tests. Still, we can certainly construct theories that are compatible with the data that we have, and evaluate them in comparison to each other. Even in the absence of intersubjective observation, there are numerous criteria available for the evaluation of such theories: simplicity, internal coherence, coherence with theories in other domains, the ability to reproduce the properties of experience that are familiar from our own case, and even an overall fit with the dictates of common sense. Perhaps there will be significant indeterminacies remaining even

when all these constraints are applied, but we can at least develop plausible candidates. Only when candidate theories have been developed will we be able to evaluate them.

A nonreductive theory of consciousness will consist in a number of *psychophysical principles*, principles connecting the properties of physical processes to the properties of experience. We can think of these principles as encapsulating the way in which experience arises from the physical. Ultimately, these principles should tell us what sort of physical systems will have associated experiences, and for the systems that do, they should tell us what sort of physical properties are relevant to the emergence of experience, and just what sort of experience we should expect any given physical system to yield. This is a tall order, but there is no reason why we should not get started.

In what follows, I present my own candidates for the psychophysical principles that might go into a theory of consciousness. The first two of these are *nonbasic principles* – systematic connections between processing and experience at a relatively high level. These principles can play a significant role in developing and constraining a theory of consciousness, but they are not cast at a sufficiently fundamental level to qualify as truly basic laws. The final principle is my candidate for a *basic principle* that might form the cornerstone of a fundamental theory of consciousness. This final principle is particularly speculative, but it is the kind of speculation that is required if we are ever to have a satisfying theory of consciousness. I can present these principles only briefly here; I argue for them at much greater length in Chalmers (1996).

The Principle of Structural Coherence

This is a principle of coherence between the *structure of consciousness* and the *structure of awareness*. Recall that “awareness” was used earlier to refer to the various functional phenomena that are associated with consciousness. I am now using it to refer to a somewhat more specific process in the cognitive underpinnings of experience. In particular, the contents of awareness are to be understood as those information contents that are accessible to central systems, and brought to bear in a widespread way in the control of behavior. Briefly put, we can think of awareness as *direct availability for global control*. To a first approximation, the contents of awareness are the contents that are directly accessible and potentially reportable, at least in a language-using system.

Awareness is a purely functional notion, but it is nevertheless intimately linked to conscious experience. In familiar cases, wherever we find consciousness, we find awareness. Wherever there is conscious experience, there is some corresponding information in the cognitive system that is available in the control of behavior, and available for verbal report. Conversely, it seems that whenever information is available for report and for global control, there is a corresponding conscious experience. Thus, there is a direct correspondence between consciousness and awareness.

The correspondence can be taken further. It is a central fact about experience that it has a complex structure. The visual field has a complex geometry, for instance. There are also relations of similarity and difference between experiences, and relations in such things as relative intensity. Every subject’s experience can be at least partly characterized and decomposed in terms of these structural properties: similarity and difference relations, perceived location, relative intensity, geometric structure, and so on. It is also a central fact that to each of these structural features, there is a corresponding feature in the information-processing structure of awareness.