

# *Can We Solve the Mind–Body Problem?*

COLIN MCGINN

How it is that anything so remarkable as a state of consciousness comes about as a result of initiating nerve tissue, is just as unaccountable as the appearance of the Djinn, where Aladdin rubbed his lamp in the story . . . (Julian Huxley)

We have been trying for a long time to solve the mind–body problem. It has stubbornly resisted our best efforts. The mystery persists. I think the time has come to admit candidly that we cannot resolve the mystery. But I also think that this very insolubility—or the reason for it—removes the philosophical problem. In this paper I explain why I say these outrageous things.

The specific problem I want to discuss concerns consciousness, the hard nut of the mind–body problem. How is it possible for conscious states to depend upon brain states? How can technicolour phenomenology arise from soggy grey matter? What makes the bodily organ we call the brain so radically different from other bodily organs, say the kidneys—the body parts without a trace of consciousness? How could the aggregation of millions of individually insentient neurons generate subjective awareness? We know that brains are the *de facto* causal basis of consciousness, but we have, it seems, no understanding whatever of how this can be so. It strikes us as miraculous, eerie, even faintly comic. Somehow, we feel, the water of the physical brain is turned into the wine of consciousness, but we draw a total blank on the nature of this conversion. Neural transmissions just seem like the wrong kind of materials with which to bring consciousness into the world, but it appears that in some way they perform this mysterious feat. The mind–body problem is the problem of understanding how the miracle is wrought, thus removing the sense of deep mystery. We want to take the magic out of the link between consciousness and the brain.<sup>1</sup>

Purported solutions to the problem have tended to assume one of two

<sup>1</sup> One of the peculiarities of the mind–body problem is the difficulty of formulating it in a rigorous way. We have a sense of the problem that outruns our capacity to articulate it clearly. Thus we quickly find ourselves resorting to invitations to look inward, instead of specifying precisely *what* it is about consciousness that makes it inexplicable in terms of ordinary physical properties. And this can make it seem that the problem is spurious. A creature without consciousness would not properly appreciate the problem (assuming such a creature could appreciate other problems). I think an adequate treatment of the mind–body problem should explain why it is so hard to state the problem explicitly. My treatment locates our difficulty in our inadequate conceptions of the nature of the brain and consciousness. In fact, if we knew their natures fully we would already have solved the problem. This should become clear later.

forms. One form, which we may call constructive, attempts to specify some natural property of the brain (or body) which explains how consciousness can be elicited from it. Thus functionalism, for example, suggests a property—namely, causal role—which is held to be satisfied by both brain states and mental states; this property is supposed to explain how conscious states can come from brain states.<sup>2</sup> The other form, which has been historically dominant, frankly admits that nothing merely natural could do the job, and suggests instead that we invoke supernatural entities or divine interventions. Thus we have Cartesian dualism and Leibnizian pre-established harmony. These ‘solutions’ at least recognize that something pretty remarkable is needed if the mind–body relation is to be made sense of; they are as extreme as the problem. The approach I favour is naturalistic but not constructive: I do not believe we can ever specify what it is about the brain that is responsible for consciousness, but I am sure that whatever it is it is not inherently miraculous. The problem arises, I want to suggest, because we are cut off by our very cognitive constitution from achieving a conception of that natural property of the brain (or of consciousness) that accounts for the psychophysical link. This is a kind of causal nexus that we are precluded from ever understanding, given the way we have to form our concepts and develop our theories. No wonder we find the problem so difficult!

Before I can hope to make this view plausible, I need to sketch the general conception of cognitive competence that underlies my position. Let me introduce the idea of *cognitive closure*. A type of mind *M* is cognitively closed with respect to a property *P* (or theory *T*) if and only if the concept-forming procedures at *M*'s disposal cannot extend to a grasp of *P* (or an understanding of *T*). Conceiving minds come in different kinds, equipped with varying powers and limitations, biases and blindspots, so that properties (or theories) may be accessible to some minds but not to others. What is closed to the mind of a rat may be open to the mind of a monkey, and what is open to us may be closed to the monkey. Representational power is not all or nothing. Minds are biological products like bodies, and like bodies they come in different shapes and sizes, more or less capacious, more or less suited to certain cognitive tasks.<sup>3</sup> This is particularly clear for perceptual faculties, of

<sup>2</sup> I would also classify panpsychism as a constructive solution, since it attempts to explain consciousness in terms of properties of the brain that are as natural as consciousness itself. Attributing specks of proto-consciousness to the constituents of matter is not supernatural in the way postulating immaterial substances or divine interventions is; it is merely extravagant. I shall here be assuming that panpsychism, like all other extant constructive solutions, is inadequate as an answer to the mind–body problem—as (of course) are the supernatural ‘solutions’. I am speaking to those who still feel perplexed (almost everyone, I would think, at least in their heart).

<sup>3</sup> This kind of view of cognitive capacity is forcefully advocated by Noam Chomsky in *Reflections on Language*, Pantheon Books, 1975, and by Jerry Fodor in *The Modularity of Mind*, Cambridge, Mass., MIT Press, 1983. Chomsky distinguishes between ‘problems’, which human minds are in principle equipped to solve, and ‘mysteries’, which systematically elude our understanding, and he envisages a study of our cognitive systems that would chart these powers and limitations. I am here engaged in such a study, citing the mind–body problem as falling on the side of the mysteries.

course: perceptual closure is hardly to be denied. Different species are capable of perceiving different properties of the world, and no species can perceive every property things may instantiate (without artificial instrumentation anyway). But such closure does not reflect adversely on the reality of the properties that lie outside the representational capacities in question; a property is no less real for not being reachable from a certain kind of perceiving and conceiving mind. The invisible parts of the electromagnetic spectrum are just as real as the visible parts, and whether a specific kind of creature can form conceptual representations of these imperceptible parts does not determine whether they exist. Thus cognitive closure with respect to *P* does not imply irrealism about *P*. That *P* is (as we might say) *noumenal* for *M* does not show that *P* does not occur in some naturalistic scientific theory *T*—it shows only that *T* is not cognitively accessible to *M*. Presumably monkey minds and the property of being an electron illustrate this possibility. And the question must arise as to whether human minds are closed with respect to certain true explanatory theories. Nothing, at least, in the concept of reality shows that everything real is open to the human concept-forming faculty—if, that is, we are realists about reality.<sup>4</sup>

Consider a mind constructed according to the principles of classical empiricism, a Humean mind. Hume mistakenly thought that human minds were Humean, but we can at least conceive of such a mind (perhaps dogs and monkeys have Humean minds). A Humean mind is such that perceptual closure determines cognitive closure, since ‘ideas’ must always be copies of ‘impressions’; therefore the concept-forming system cannot transcend what can be perceptually presented to the subject. Such a mind will be closed with respect to unobservables; the properties of atoms, say, will not be representable by a mind constructed in this way. This implies that explanatory theories in which these properties are essentially mentioned will not be accessible to a Humean mind.<sup>5</sup> And hence the observable phenomena that are explained by allusion to unobservables will be inexplicable by a mind thus limited. But notice: the incapacity to explain certain phenomena does not carry with it a lack of recognition of the theoretical problems the phenomena pose. You might be able to

<sup>4</sup> See Thomas Nagel’s discussion of realism in *The View From Nowhere*, Oxford, Oxford University Press, 1986, ch. VI. He argues there for the possibility of properties we can never grasp. Combining Nagel’s realism with Chomsky–Fodor cognitive closure gives a position looking very much like Locke’s in the *Essay Concerning Human Understanding*: the idea that our God-given faculties do not equip us to fathom the deep truth about reality. In fact, Locke held precisely this about the relation between mind and brain: only divine revelation could enable us to understand how ‘perceptions’ are produced in our minds by material objects.

<sup>5</sup> Hume, of course, argued, in effect, that no theory essentially employing a notion of objective causal necessitation could be grasped by our minds—and likewise for the notion of objective persistence. We might compare the frustrations of the Humean mind to the conceptual travails of the pure sound beings discussed in Ch. II of P. F. Strawson’s *Individuals*, London, Methuen, 1959; both are types of mind whose constitution puts various concepts beyond them. We can do a lot better than these truncated minds, but we also have our constitutional limitations.

appreciate a problem without being able to formulate (even in principle) the solution to that problem (I suppose human children are often in this position, at least for a while). A Humean mind cannot solve the problems that our physics solves, yet it might be able to have an inkling of what needs to be explained. We would expect, then, that a moderately intelligent enquiring Humean mind will feel permanently perplexed and mystified by the physical world, since the correct science is forever beyond its cognitive reach. Indeed, something like this was precisely the view of Locke. He thought that our ideas of matter are quite sharply constrained by our perceptions and so concluded that the true science of matter is eternally beyond us—that we could never remove our perplexities about (say) what solidity ultimately is.<sup>6</sup> But it does not follow for Locke that nature is itself inherently mysterious; the felt mystery comes from our own cognitive limitations, not from any objective eeriness in the world. It looks today as if Locke was wrong about our capacity to fathom the nature of the physical world, but we can still learn from his fundamental thought—the insistence that our cognitive faculties may not be up to solving every problem that confronts us. To put the point more generally: the human mind may not conform to empiricist principles, but it must conform to *some* principles—and it is a substantive claim that these principles permit the solution of every problem we can formulate or sense. Total cognitive openness is not guaranteed for human beings and it should not be expected. Yet what is noumenal for us may not be miraculous in itself. We should therefore be alert to the possibility that a problem that strikes us as deeply intractable, as utterly baffling, may arise from an area of cognitive closure in our ways of representing the world.<sup>7</sup> That is what I now want to argue is the case with our sense of the mysterious nature of the connection between consciousness and the brain. We are biased away from arriving at the correct explanatory theory of the psychophysical nexus. And this makes us prone to an illusion of objective mystery. Appreciating this should remove the philosophical problem: consciousness does not, in reality, arise from the brain in the miraculous way in which the Djin arises from the lamp.

I now need to establish three things: (i) there exists some property of the brain that accounts naturalistically for consciousness; (ii) we are cognitively closed with respect to that property; but (iii) there is no philosophical (as opposed to scientific) mind–body problem. Most of the work will go into establishing (ii).

<sup>6</sup> See the *Essay*, Book II, ch. IV. Locke compares the project of saying what solidity ultimately is to trying to clear up a blind man's vision by talking to him.

<sup>7</sup> Some of the more arcane aspects of cosmology and quantum theory might be thought to lie just within the bounds of human intelligibility. Chomsky suggests that the causation of behaviour might be necessarily mysterious to human investigators: see *Reflections on Language*, p. 156. I myself believe that the mind–body problem exhibits a qualitatively different level of mystery from this case (unless it is taken as an aspect of that problem).

Resolutely shunning the supernatural, I think it is undeniable that it must be in virtue of *some* natural property of the brain that organisms are conscious. There just *has* to be some explanation for how brains subserve minds. If we are not to be eliminativists about consciousness, then some theory must exist which accounts for the psychophysical correlations we observe. It is implausible to take these correlations as ultimate and inexplicable facts, as simply brute. And we do not want to acknowledge radical emergence of the conscious with respect to the cerebral: that is too much like accepting miracles *de re*. Brain states cause conscious states, we know, and this causal nexus must proceed through necessary connections of some kind—the kind that would make the nexus intelligible *if* they were understood.<sup>8</sup> Consciousness is like life in this respect. We know that life evolved from inorganic matter, so we expect there to be some explanation of this process. We cannot plausibly take the arrival of life as a primitive brute fact, nor can we accept that life arose by some form of miraculous emergence. Rather, there must be some natural account of how life comes from matter, whether or not we can know it. Eschewing vitalism and the magic touch of God's finger, we rightly insist that it must be in virtue of some natural property of (organized) matter that parcels of it get to be alive. But consciousness itself is just a further biological development, and so it too must be susceptible of some natural explanation—whether or not human beings are capable of arriving at this explanation. Presumably there exist objective natural laws that somehow account for the upsurge of consciousness. Consciousness, in short, must be a natural phenomenon, naturally arising from certain organizations of matter. Let us then say that there exists some property *P*, instantiated by the brain, in virtue of which the brain is the basis of consciousness. Equivalently, there exists some theory *T*, referring to *P*, which fully explains the dependence of conscious states on brain states. If we knew *T*, then we would have a constructive solution to the mind–body problem. The question then is whether we can ever come to know *T* and grasp the nature of *P*.

Let me first observe that it is surely *possible* that we could never arrive at a grasp of *P*; there is, as I said, no guarantee that our cognitive powers permit the solution of every problem we can recognize. Only a misplaced idealism about the natural world could warrant the dogmatic claim that everything is knowable by the human species at this stage of its evolutionary development (consider the same claim made on behalf of the intellect of cro-Magnon man). It *may* be that every property for which we can form a concept is such that *it* could never solve the mind–body problem. We *could* be like five-year old children trying to understand

<sup>8</sup> Cf. Nagel's discussion of emergence in 'Panpsychism', in *Mortal Questions*, Cambridge, Cambridge University Press, 1979. I agree with him that the apparent radical emergence of mind from matter has to be epistemic only, on pain of accepting inexplicable miracles in the world.

Relativity Theory. Still, so far this is just a possibility claim: what reason do we have for asserting, positively, that our minds are closed with respect to *P*?

Longstanding historical failure is suggestive, but scarcely conclusive. Maybe, it will be said, the solution is just around the corner, or it has to wait upon the completion of the physical sciences? Perhaps we simply have yet to produce the Einstein-like genius who will restructure the problem in some clever way and then present an astonished world with the solution?<sup>9</sup> However, I think that our deep bafflement about the problem, amounting to a vertiginous sense of ultimate mystery, which resists even articulate formulation, should at least encourage us to explore the idea that there is something terminal about our perplexity. Rather as traditional theologians found themselves conceding cognitive closure with respect to certain of the properties of God, so we should look seriously at the idea that the mind–body problem brings us bang up against the limits of our capacity to understand the world. That is what I shall do now.

There seem to be two possible avenues open to us in our aspiration to identify *P*: we could try to get to *P* by investigating consciousness directly, or we could look to the study of the brain for *P*. Let us consider these in turn, starting with consciousness. Our acquaintance with consciousness could hardly be more direct; phenomenological description thus comes (relatively) easily. ‘Introspection’ is the name of the faculty through which we catch consciousness in all its vivid nakedness. By virtue of possessing this cognitive faculty we ascribe concepts of consciousness to ourselves; we thus have ‘immediate access’ to the properties of consciousness. But does the introspective faculty reveal property *P*? Can we tell just by introspecting what the solution to the mind–body problem is? Clearly not. We have direct cognitive access to one term of the mind–brain relation, but we do not have such access to the nature of the link. Introspection does not present conscious states *as* depending upon the brain in some intelligible way. We cannot therefore introspect *P*. Moreover, it seems impossible that we should ever augment our stock of introspectively ascribed concepts with the concept *P*—that is, we could not acquire this concept simply on the basis of sustained and careful introspection. Pure phenomenology will never provide the solution to the mind–body problem. Neither does it seem feasible to try to extract *P* from the concepts of consciousness we now have by some procedure of conceptual analysis—any more than we could solve the life–matter problem simply by reflecting on the concept

<sup>9</sup> Despite his reputation for pessimism over the mind–body problem, a careful reading of Nagel reveals an optimistic strain in his thought (by the standards of the present paper): see, in particular, the closing remarks of ‘What is it Like to be a Bat?’, in *Mortal Questions*. Nagel speculates that we might be able to devise an ‘objective phenomenology’ that made conscious states more amenable to physical analysis. Unlike me, he does not regard the problem as inherently beyond us.

life.<sup>10</sup> *P* has to lie outside the field of the introspectable, and it is not implicitly contained in the concepts we bring to bear in our first-person ascriptions. Thus the faculty of introspection, as a concept-forming capacity, is cognitively closed with respect to *P*; which is not surprising in view of its highly limited domain of operation (*most* properties of the world are closed to introspection).

But there is a further point to be made about *P* and consciousness, which concerns our restricted access to the concepts of consciousness themselves. It is a familiar point that the range of concepts of consciousness attainable by a mind *M* is constrained by the specific forms of consciousness possessed by *M*. Crudely, you cannot form concepts of conscious properties unless you yourself instantiate those properties. The man born blind cannot grasp the concept of a visual experience of red, and human beings cannot conceive of the echolocatory experiences of bats.<sup>11</sup> These are cases of cognitive closure within the class of conscious properties. But now this kind of closure will, it seems, affect our hopes of access to *P*. For suppose that we were cognitively open with respect to *P*; suppose, that is, that we had the solution to the problem of how specific forms of consciousness depend upon different kinds of physiological structure. Then, of course, we would understand how the brain of a bat subserves the subjective experiences of bats. Call this type of experience *B*, and call the explanatory property that links *B* to the bat's brain *P*<sub>1</sub>. By grasping *P*<sub>1</sub> it would be perfectly intelligible to us how the bat's brain generates *B*-experiences; we would have an explanatory theory of the causal nexus in question. We would be in possession of the same kind of understanding we would have of our own experiences if we had the correct psychophysical theory of them. But then it seems to follow that grasp of the theory that explains *B*-experiences would *confer* a grasp of the nature of those experiences: for how could we understand that theory without understanding the concept *B* that occurs in it? How could we grasp the *nature* of *B*-experiences without grasping the *character* of those experiences? The true psychophysical theory would seem to provide a route to a grasp of the subjective form of the bat's experiences. But now we face a dilemma, a dilemma which threatens to become a *reductio*: either we *can* grasp this theory, in which case the property *B* becomes open to us; or we

<sup>10</sup> This is perhaps the most remarkably optimistic view of all—the expectation that reflecting on the ordinary concept of pain (say) will reveal the manner of pain's dependence on the brain. If I am not mistaken, this is in effect the view of common-sense functionalists: they think that *P* consists in causal role, and that this can be inferred analytically from the concepts of conscious states. This would make it truly amazing that we should ever have felt there to be a mind–body problem at all, since the solution is already contained in our mental concepts. What optimism!

<sup>11</sup> See Nagel, 'What is it Like to be a Bat?' Notice that the fugitive character of such properties with respect to our concepts has nothing to do with their 'complexity'; like fugitive colour properties, such experiential properties are 'simple'. Note too that such properties provide counter-examples to the claim that (somehow) rationality is a faculty that, once possessed, can be extended to encompass all concepts, so that if *any* concept can be possessed then *every* concept can.

cannot grasp the theory, simply because property *B* is *not* open to us. It seems to me that the looming reductio here is compelling: our concepts of consciousness just *are* inherently constrained by our own form of consciousness, so that any theory the understanding of which required us to transcend these constraints would *ipso facto* be inaccessible to us. Similarly, I think, any theory that required us to transcend the finiteness of our cognitive capacities would *ipso facto* be a theory we could not grasp—and this despite the fact that it might be needed to explain something we can see needs explaining. We cannot simply stipulate that our concept-forming abilities are indefinitely plastic and unlimited just because they would have to be to enable us to grasp the truth about the world. We constitutionally lack the concept-forming capacity to encompass all possible types of conscious state, and this obstructs our path to a general solution to the mind–body problem. Even if we could solve it for our own case, we could not solve it for bats and Martians. *P* is, as it were, too close to the different forms of subjectivity for it to be accessible to all such forms, given that one’s form of subjectivity restricts one’s concepts of subjectivity.<sup>12</sup>

I suspect that most optimists about constructively solving the mind–body problem will prefer to place their bets on the brain side of the relation. Neuroscience is the place to look for property *P*, they will say. My question then is whether there is any conceivable way in which we might come to introduce *P* in the course of our empirical investigations of the brain. New concepts have been introduced in the effort to understand the workings of the brain, certainly: could not *P* then occur in conceivable extensions of this manner of introduction? So far, indeed, the theoretical concepts we ascribe to the brain seem as remote from consciousness as any ordinary physical properties are, but perhaps we might reach *P* by diligent application of essentially the same procedures: so it is tempting to think. I want to suggest, to the contrary, that such procedures are inherently closed with respect to *P*. The fundamental reason for this, I think, is the role of *perception* in shaping our understanding of the brain—the way that our perception of the brain constrains the concepts we can apply to it. A point whose significance it would be hard to overstress here is this: the property of consciousness itself (or specific conscious states) is not an observable or

<sup>12</sup> It might be suggested that we borrow Nagel’s idea of ‘objective phenomenology’ in order to get around this problem. Instead of representing experiences under subjective descriptions, we should describe them in entirely objective terms, thus bringing them within our conceptual ken. My problem with this is that, even allowing that there could be such a form of description, it would not permit us to understand how the subjective aspects of experience depend upon the brain—which is really the problem we are trying to solve. In fact, I doubt that the notion of objective phenomenology is any more coherent than the notion of subjective physiology. Both involve trying to bridge the psychophysical gap by a sort of stipulation. The lesson here is that the gap cannot be bridged just by applying concepts drawn from one side to items that belong on the other side; and this is because neither sort of concept could ever do what is needed.

perceptible property of the brain. You can stare into a living conscious brain, your own or someone else's, and see there a wide variety of unstantiated properties—its shape, colour, texture, etc.—but you will not thereby *see* what the subject is experiencing, the conscious state itself. Conscious states are simply not potential objects of perception: they depend upon the brain but they cannot be observed by directing the senses onto the brain. In other words, consciousness is noumenal with respect to perception of the brain.<sup>13</sup> I take it this is obvious. So we know there *are* properties of the brain that are necessarily closed to perception of the brain; the question now is whether *P* is likewise closed to perception.

My argument will proceed as follows. I shall first argue that *P* is indeed perceptually closed; then I shall complete the argument to full cognitive closure by insisting that no form of *inference* from what is perceived can lead us to *P*. The argument for perceptual closure starts from the thought that nothing we can imagine perceiving in the brain would ever convince us that we have located the intelligible nexus we seek. No matter what recondite property we could see to be instantiated in the brain we would always be baffled about how it could give rise to consciousness. I hereby invite you to try to conceive of a perceptible property of the brain that might allay the feeling of mystery that attends our contemplation of the brain–mind link: I do not think you will be able to do it. It is like trying to conceive of a perceptible property of a rock that would render it perspicuous that the rock was conscious. In fact, I think it is the very impossibility of this that lies at the root of the felt mind–body problem. But why is this? Basically, I think, it is because the senses are geared to representing a spatial world; they essentially present things in space with spatially defined properties. But it is precisely *such* properties that seem inherently incapable of resolving the mind–body problem: we cannot link consciousness to the brain in virtue of spatial properties of the brain. There the brain is, an object of perception, laid out in space, containing spatially distributed processes; but consciousness defies explanation in such terms. Consciousness does not seem made up out of smaller spatial processes; yet perception of the brain seems limited to revealing such processes.<sup>14</sup> The senses are responsive to certain *kinds* of properties—those that are essentially bound up with space—but these

<sup>13</sup> We should distinguish two claims about the imperceptibility of consciousness: (i) consciousness is not perceivable by directing the senses onto the brain; (ii) consciousness is not perceivable by directing the senses anywhere, even towards the behaviour that 'expresses' conscious states. I believe both theses, but my present point requires only (i). I am assuming, of course, that perception cannot be unrestrictedly theory-laden; or that if it can, the infusions of theory cannot have been originally derived simply by looking at things or tasting them or touching them or . . .

<sup>14</sup> Nagel discusses the difficulty of thinking of conscious processes in the spatial terms that apply to the brain in *The View From Nowhere*, pp. 50–1, but he does not draw my despairing conclusion. The case is exactly *unlike* (say) the dependence of liquidity on the properties of molecules, since here we do think of both terms of the relation as spatial in character; so we can simply employ the idea of spatial composition.

properties are of the wrong sort (the wrong *category*) to constitute *P*. Kant was right, the form of outer sensibility is spatial; but if so, then *P* will be noumenal with respect to the senses, since no spatial property will ever deliver a satisfying answer to the mind–body problem. We simply do not understand the idea that conscious states might intelligibly arise from spatial configurations of the kind disclosed by perception of the world.

I take it this claim will not seem terribly controversial. After all, we do not generally expect that every property referred to in our theories should be a potential object of human perception: consider quantum theory and cosmology. Unrestricted perceptual openness is a dogma of empiricism if ever there was one. And there is no compelling reason to suppose that the property needed to explain the mind–brain relation should be in principle perceptible; it might be essentially ‘theoretical’, an object of thought not sensory experience. Looking harder at nature is not the only (or the best) way of discovering its theoretically significant properties. Perceptual closure does not entail cognitive closure, since we have available the procedure of hypothesis formation, in which *unobservables* come to be conceptualized.

I readily agree with these sentiments, but I think there are reasons for believing that no coherent method of concept introduction will ever lead us to *P*. This is because a certain principle of *homogeneity* operates in our introduction of theoretical concepts on the basis of observation. Let me first note that consciousness itself could not be introduced simply on the basis of what we observe about the brain and its physical effects. If our data, arrived at by perception of the brain, do not include anything that brings in conscious states, then the theoretical properties we need to explain these data will not include conscious states either. Inference to the best explanation of purely physical data will never take us outside the realm of the physical, forcing us to introduce concepts of consciousness.<sup>15</sup> Everything physical has a purely physical explanation. So the property of consciousness is cognitively closed with respect to the introduction of concepts by means of inference to the best explanation of perceptual data about the brain.

Now the question is whether *P* could ever be arrived at by this kind of inference. Here we must be careful to guard against a form of magical emergentism with respect to concept formation. Suppose we try out a relatively clear theory of how theoretical concepts are formed: we get them by a sort of analogical extension of what we observe. Thus, for example, we arrive at the concept of a molecule by taking our perceptual representations of macroscopic objects and conceiving of smaller scale objects of the same general kind. This method seems to work well enough

<sup>15</sup> Cf. Nagel: ‘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’, ‘Panpsychism’, p. 183.

for unobservable material objects, but it will not help in arriving at *P*, since analogical extensions of the entities we observe in the brain are precisely as hopeless as the original entities were as solutions to the mind–body problem. We would need a method that left the base of observational properties behind in a much more radical way. But it seems to me that even a more unconstrained conception of inference to the best explanation would still not do what is required: it would no more serve to introduce *P* than it serves to introduce the property of consciousness itself. To explain the observed physical data we need only such theoretical properties as bear upon those data, not the property that explains consciousness, which does not occur in the data. Since we do not need consciousness to explain those data, we do not need the property that explains consciousness. We will never get as far away from the perceptual data in our explanations of those data as we need to get in order to connect up explanatorily with consciousness. This is, indeed, why it seems that consciousness is theoretically epiphenomenal in the task of accounting for physical events. No concept needed to explain the workings of the physical world will suffice to explain how the physical world produces consciousness. So if *P* is perceptually noumenal, then it will be noumenal with respect to perception-based explanatory inferences. Accordingly, I do not think that *P* could be arrived at by empirical studies of the brain alone. Nevertheless, the brain *has* this property, as it has the property of consciousness. Only a magical idea of how we come by concepts could lead one to think that we can reach *P* by first perceiving the brain and then asking what is needed to explain what we perceive.<sup>16</sup> (The mind–body problem tempts us to magic in more ways than one.)

It will help elucidate the position I am driving towards if I contrast it with another view of the source of the perplexity we feel about the mind–brain nexus. I have argued that we cannot know which property of the brain accounts for consciousness, and so we find the mind–brain link unintelligible. But, it may be said, there is another account of our sense of irremediable mystery, which does not require positing properties our minds cannot represent. This alternative view claims that, even if we *now* had a grasp of *P*, we would *still* feel that there is something mysterious about the link, because of a special epistemological feature of the situation. Namely this: our acquaintance with the brain and our acquaintance with consciousness are necessarily mediated by distinct cognitive faculties, namely perception and introspection. Thus the faculty through which we

<sup>16</sup> It is surely a striking fact that the microprocesses that have been discovered in the brain by the usual methods seem no nearer to consciousness than the gross properties of the brain open to casual inspection. Neither do more abstract 'holistic' features of brain function seem to be on the right lines to tell us the nature of consciousness. The deeper science probes into the brain the more remote it seems to get from consciousness. Greater knowledge of the brain thus destroys our illusions about the kinds of properties that might be discovered by travelling along this path. Advanced neurophysiological theory seems only to deepen the miracle.

apprehend one term of the relation is necessarily distinct from the faculty through which we apprehend the other. In consequence, it is not possible for us to use one of these faculties to apprehend the nature of the psychophysical nexus. No single faculty will enable us ever to apprehend the fact that consciousness depends upon the brain in virtue of property *P*. Neither perception alone nor introspection alone will ever enable us to witness the dependence. And this, my objector insists, is the real reason we find the link baffling: we cannot make sense of it in terms of the deliverances of a single cognitive faculty. So, even if we now had concepts for the properties of the brain that explain consciousness, we would still feel a residual sense of unintelligibility; we would still take there to be something mysterious going on. The necessity to shift from one faculty to the other produces in us an illusion of inexplicability. We might in fact have the explanation right now but be under the illusion that we do not. The right diagnosis, then, is that we should recognize the peculiarity of the epistemological situation and stop trying to make sense of the psychophysical nexus in the way we make sense of other sorts of nexus. It only *seems* to us that we can never discover a property that will render the nexus intelligible.

I think this line of thought deserves to be taken seriously, but I doubt that it correctly diagnoses our predicament. It is true enough that the problematic nexus is essentially apprehended by distinct faculties, so that it will never reveal its secrets to a single faculty; but I doubt that our intuitive sense of intelligibility is so rigidly governed by the 'single-faculty condition'. Why *should* facts only seem intelligible to us if we can conceive of apprehending them by one (sort of) cognitive faculty? Why not allow that we can recognize intelligible connections between concepts (or properties) even when those concepts (or properties) are necessarily ascribed using different faculties? Is it not suspiciously empiricist to insist that a causal nexus can only be made sense of by us if we can conceive of its being an object of a single faculty of apprehension? Would we think this of a nexus that called for touch and sight to apprehend each term of the relation? Suppose (*per impossibile*) that we were offered *P* on a plate, as a gift from God: would we still shake our heads and wonder how that could resolve the mystery, being still the victims of the illusion of mystery generated by the epistemological duality in question? No, I think this suggestion is not enough to account for the miraculous appearance of the link: it is better to suppose that we are permanently blocked from forming a concept of what accounts for that link.

How strong is the thesis I am urging? Let me distinguish *absolute* from *relative* claims of cognitive closure. A problem is absolutely cognitively closed if no possible mind could resolve it; a problem is relatively closed if minds of some sorts can in principle solve it while minds of other sorts cannot. Most problems we may safely suppose, are only relatively closed:

armadillo minds cannot solve problems of elementary arithmetic but human minds can. Should we say that the mind–body problem is only relatively closed or is the closure absolute? This depends on what we allow as a possible concept-forming mind, which is not an easy question. If we allow for minds that form their concepts of the brain and consciousness in ways that are quite independent of perception and introspection, then there may be room for the idea that there are possible minds for which the mind–body problem is soluble, and easily so. But if we suppose that *all* concept formation is tied to perception and introspection, however loosely, then *no* mind will be capable of understanding how it relates to its own body—the insolubility will be absolute. I think we can just about make sense of the former kind of mind, by exploiting our own faculty of a priori reasoning. Our mathematical concepts (say) do not seem tied either to perception or to introspection, so there does seem to be a mode of concept formation that operates without the constraints I identified earlier. The suggestion might then be that a mind that formed all of its concepts in this way—including its concepts of the brain and consciousness—would be free of the biases that prevent *us* from coming up with the right theory of how the two connect. Such a mind would have to be able to think of the brain and consciousness in ways that utterly prescind from the perceptual and the introspective—in somewhat the way we now (it seems) think about numbers. This mind would conceive of the psychophysical link in totally a priori terms. Perhaps this is how we should think of God’s mind, and God’s understanding of the mind–body relation. At any rate, something pretty radical is going to be needed if we are to devise a mind that can escape the kinds of closure that make the problem insoluble for us—if I am right in my diagnosis of our difficulty. *If* the problem is only relatively insoluble, then the type of mind that can solve it is going to be very different from ours and the kinds of mind we can readily make sense of (there may, of course, be cognitive closure here too). It certainly seems to me to be at least an open question whether the problem is absolutely insoluble; I would not be surprised if it were.<sup>17</sup>

My position is both pessimistic and optimistic at the same time. It is pessimistic about the prospects for arriving at a constructive solution to the mind–body problem, but it is optimistic about our hopes of removing the philosophical perplexity. The central point here is that I do not think we need to do the former in order to achieve the latter. This depends on a rather special understanding of what the philosophical problem consists in. What I want to suggest is that the nature of the psychophysical connection has a full and non-mysterious explanation in a certain science, but that this

<sup>17</sup> The kind of limitation I have identified is therefore not the kind that could be remedied simply by a large increase in general intelligence. No matter how large the frontal lobes of our biological descendants may become, they will still be stumped by the mind–body problem, so long as they form their (empirical) concepts on the basis of perception and introspection.

science is inaccessible to us as a matter of principle. Call this explanatory scientific theory *T*: *T* is as natural and prosaic and devoid of miracle as any theory of nature; it describes the link between consciousness and the brain in a way that is no more remarkable (or alarming) than the way we now describe the link between the liver and bile.<sup>18</sup> According to *T*, there is nothing eerie going on in the world when an event in my visual cortex causes me to have an experience of yellow—however much it seems to *us* that there is. In other words, there is no intrinsic conceptual or metaphysical difficulty about how consciousness depends on the brain. It is not that the correct science is compelled to postulate miracles *de re*; it is rather that the correct science lies in the dark part of the world for us. We confuse our own cognitive limitations with objective eeriness. We are like a Humean mind trying to understand the physical world, or a creature without spatial concepts trying to understand the possibility of motion. This removes the philosophical problem because it assures us that the entities *themselves* pose no inherent philosophical difficulty. The case is unlike, for example, the problem of how the abstract world of numbers might be intelligibly related to the world of concrete knowing subjects: here the mystery seems intrinsic to the entities, not a mere artefact of our cognitive limitations or biases in trying to understand the relation.<sup>19</sup> It would not be plausible to suggest that there exists a science, whose theoretical concepts we cannot grasp, which completely resolves any sense of mystery that surrounds the question how the abstract becomes an object of knowledge for us. In this case, then, eliminativism seems a live option. The *philosophical* problem about consciousness and the brain arises from a sense that we are compelled to accept that nature contains miracles—as if the merely metallic lamp of the brain could really spirit into existence the Djinn of consciousness. But we do not need to accept this: we can rest secure in the knowledge that some (unknowable) property of the brain makes everything fall into place. What creates the philosophical puzzle is the assumption that the problem must somehow be scientific but

<sup>18</sup> Or again, no more miraculous than the theory of evolution. Creationism is an understandable response to the theoretical problem posed by the existence of complex organisms; fortunately, we now have a theory that renders this response unnecessary, and so undermines the theism required by the creationist thesis. In the case of consciousness, the appearance of miracle might also tempt us in a 'creationist' direction, with God required to perform the alchemy necessary to transform matter into experience. Thus the mind-body problem might similarly be used to prove the existence of God (no miracle without a miracle-maker). We cannot, I think, refute this argument in the way we can the original creationist argument, namely by actually producing a non-miraculous explanatory theory, but we can refute it by arguing that such a naturalistic theory must *exist*. (It is a condition of adequacy upon any account of the mind-body relation that it avoid assuming theism.)

<sup>19</sup> See Paul Benacerraf, 'Mathematical Truth', *Journal of Philosophy*, 1973, for a statement of this problem about abstract entities. Another problem that seems to me to differ from the mind-body problem is the problem of free will. I do not believe that there is some unknowable property *Q* which reconciles free will with determinism (or indeterminism); rather, the concept of free will contains internal incoherencies—as the concept of consciousness does not. This is why it is much more reasonable to be an eliminativist about free will than about consciousness.

that any science *we* can come up with will represent things as utterly miraculous. And the solution is to recognize that the sense of miracle comes from us and not from the world. There is, in reality, nothing mysterious about how the brain generates consciousness. There is no *metaphysical* problem.<sup>20</sup>

So far that deflationary claim has been justified by a general naturalism and certain considerations about cognitive closure and the illusions it can give rise to. Now I want to marshal some reasons for thinking that consciousness is actually a rather simple natural fact; objectively, consciousness is nothing very special. We should now be comfortable with the idea that our own sense of difficulty is a fallible guide to objective complexity: what is hard for us to grasp may not be very fancy in itself. The grain of our thinking is not a mirror held up to the facts of nature.<sup>21</sup> In particular, it may be that the extent of our understanding of facts about the mind is not commensurate with some objective estimate of their intrinsic complexity: we may be good at understanding the mind in some of its aspects but hopeless with respect to others, in a way that cuts across objective differences in what the aspects involve. Thus we are adept at understanding action in terms of the folk psychology of belief and desire, and we seem not entirely out of our depth when it comes to devising theories of language. But our understanding of how consciousness develops from the organization of matter is non-existent. But now, think of these various aspects of mind from the point of view of evolutionary biology. Surely language and the propositional attitudes are more complex and advanced evolutionary achievements than the mere possession of consciousness by a physical organism. Thus it seems that we are better at understanding some of the more complex aspects of mind than the simpler ones. Consciousness arises early in evolutionary history and is found right across the animal kingdom. In some respects it seems that the biological engineering required for consciousness is less fancy than that needed for certain kinds of complex motor behaviour. Yet we can come to understand the latter while drawing a total blank with respect to the former. Conscious states seem biologically quite primitive, comparatively speaking. So the

<sup>20</sup> A test of whether a proposed solution to the mind–body problem is adequate is whether it relieves the pressure towards eliminativism. If the data can only be explained by postulating a miracle (i.e. not explained), then we must repudiate the data—this is the principle behind the impulse to deny that conscious states exist. My proposal passes this test because it allows us to resist the postulation of miracles; it interprets the eeriness as merely epistemic, though deeply so. Constructive solutions are not the only way to relieve the pressure.

<sup>21</sup> Chomsky suggests that the very faculties of mind that make us good at some cognitive tasks may make us poor at others; see *Reflections on Language*, pp. 155–6. It seems to me possible that what makes us good at the science of the purely physical world is what skews us away from developing a science of consciousness. Our faculties bias us towards understanding matter in motion, but it is precisely this kind of understanding that is inapplicable to the mind–body problem. Perhaps, then, the price of being good at understanding matter is that we cannot understand mind. Certainly our notorious tendency to think of everything in spatial terms does not help us in understanding the mind.

theory *T* that explains the occurrence of consciousness in a physical world is very probably less objectively complex (by some standard) than a range of other theories that do not defy our intellects. If only we could know the psychophysical mechanism it might surprise us with its simplicity, its utter naturalness. In the manual that God consulted when he made the earth and all the beasts that dwell thereon the chapter about how to engineer consciousness from matter occurs fairly early on, well before the really difficult later chapters on mammalian reproduction and speech. It is not the *size* of the problem but its *type* that makes the mind–body problem so hard for us. This reflection should make us receptive to the idea that it is something about the tracks of our thought that prevents us from achieving a science that relates consciousness to its physical basis: the enemy lies within the gates.<sup>22</sup>

The position I have reached has implications for a tangle of intuitions it is natural to have regarding the mind–body relation. On the one hand, there are intuitions, pressed from Descartes to Kripke, to the effect that the relation between conscious states and bodily states is fundamentally contingent.<sup>23</sup> It can easily seem to us that there is no necessitation involved in the dependence of the mind on the brain. But, on the other hand, it looks absurd to try to dissociate the two entirely, to let the mind float completely free of the body. Disembodiment is a dubious possibility at best, and some kind of necessary supervenience of the mental on the physical has seemed undeniable to many. It is not my aim here to adjudicate this longstanding dispute; I want simply to offer a diagnosis of what is going on when one finds oneself assailed with this flurry of conflicting intuitions. The reason we feel the tug of contingency, pulling consciousness loose from its physical moorings, may be that we do not and cannot grasp the nature of the property that intelligibly links them. The brain has physical properties we can grasp, and variations in these correlate with changes in consciousness, but we cannot draw the veil that conceals the manner of their connection. Not grasping the nature of the connection, it strikes us as deeply contingent; we cannot make the assertion of a necessary connection intelligible to ourselves. There *may* then be a real necessary connection; it is just that it will always strike us as curiously brute and unperceptive. We may thus, as upholders of intrinsic contingency, be the dupes of our own cognitive blindness. On the other hand, we are scarcely in a position to assert that there *is* a necessary

<sup>22</sup> I get this phrase from Fodor, *The Modularity of Mind*, p. 121. The intended contrast is with kinds of cognitive closure that stem from exogenous factors—as, say, in astronomy. Our problem with *P* is not that it is too distant or too small or too large or too complex; rather, the very structure of our concept-forming apparatus points us away from *P*.

<sup>23</sup> Saul Kripke, *Naming and Necessity*, Oxford, Blackwell, 1980. Of course, Descartes explicitly argued from (what he took to be) the essential natures of the body and mind to the contingency of their connection. If we abandon the assumption that we know these natures, then agnosticism about the modality of the connection seems the indicated conclusion.

connection between the properties of the brain we can grasp and states of consciousness, since we are so ignorant (and irremediably so) about the character of the connection. For all we know, the connection may be contingent, as access to *P* would reveal if we could have such access. The link between consciousness and property *P* is not, to be sure, contingent—virtually by definition—but we are not in a position to say exactly how *P* is related to the ‘ordinary’ properties of the brain. It may be necessary or it may be contingent. Thus it is that we tend to vacillate between contingency and necessity; for we lack the conceptual resources to decide the question—or to understand the answer we are inclined to give. The indicated conclusion appears to be that we can never really know whether disembodiment is metaphysically possible, or whether necessary supervenience is the case, or whether spectrum inversion could occur. For these all involve claims about the modal connections between properties of consciousness and the ordinary properties of the body and brain that we can conceptualize; and the real nature of these connections is not accessible to us. Perhaps *P* makes the relation between C-fibre firing and pain necessary or perhaps it does not: we are simply not equipped to know. We are like a Humean mind wondering whether the observed link between the temperature of a gas and its pressure (at a constant volume) is necessary or contingent. To know the answer to that you need to grasp atomic (or molecular) theory, and a Humean mind just is not up to attaining the requisite theoretical understanding. Similarly, we are constitutionally ignorant at precisely the spot where the answer exists.

I predict that many readers of this paper will find its main thesis utterly incredible, even ludicrous. Let me remark that I sympathize with such readers: the thesis is not easily digestible. But I would say this: if the thesis is actually true, it will still strike us as hard to believe. For the idea of an explanatory property (or set of properties) that is noumenal for us, yet is essential for the (constructive) solution of a problem we face, offends a kind of natural idealism that tends to dominate our thinking. We find it taxing to conceive of the existence of a real property, under our noses as it were, which we are built not to grasp—a property that is responsible for phenomena that we observe in the most direct way possible. This kind of realism, which brings cognitive closure so close to home, is apt to seem both an affront to our intellects and impossible to get our minds around. We try to think of this unthinkable property and understandably fail in the effort; so we rush to infer that the very supposition of such a property is nonsensical. Realism of the kind I am presupposing thus seems difficult to hold in focus, and any philosophical theory that depends upon it will also seem to rest on something systematically elusive.<sup>24</sup> My response to such

<sup>24</sup> This is the kind of realism defended by Nagel in ch. VI of *The View From Nowhere*: to be is not to be conceivable by us. I would say that the mind–body problem provides a demonstration that there are such concept-transcending properties—not merely that there *could* be. I would also say that

misgivings, however, is unconcessive: the limits of our minds are just not the limits of reality. It is deplorably anthropocentric to insist that reality be constrained by what the human mind can conceive. We need to cultivate a vision of reality (a metaphysics) that makes it truly independent of our given cognitive powers, a conception that includes these powers as a proper part. It is just that, in the case of the mind–body problem, the bit of reality that systematically eludes our cognitive grasp is an aspect of our own nature. Indeed, it is an aspect that makes it possible for us to have minds at all and to think about how they are related to our bodies. This particular transcendent tract of reality happens to lie within our own heads. A deep fact about our own nature as a form of embodied consciousness is thus necessarily hidden from us. Yet there is nothing inherently eerie or bizarre about this embodiment. We are much more straightforward than we seem. Our weirdness lies in the eye of the beholder.

The answer to the question that forms my title is therefore ‘No and Yes’.<sup>25</sup>

*Oxford University*

COLIN MCGINN

realism of this kind should be accepted precisely because it helps solve the mind–body problem; it is a metaphysical thesis that pulls its weight in coping with a problem that looks hopeless otherwise. There is thus nothing ‘epiphenomenal’ about such radical realism: the existence of a reality we cannot know can yet have intellectual significance for us.

<sup>25</sup> Discussions with the following people have helped me work out the ideas of this paper: Anita Avramides, Jerry Katz, Ernie Lepore, Michael Levin, Thomas Nagel, Galen Strawson, Peter Unger. My large debt to Nagel’s work should be obvious throughout the paper: I would not have tried to face the mind–body problem down had he not first faced up to it.