

CONCEIVING THE IMPOSSIBLE AND THE MIND-BODY PROBLEM

Royal Institute of Philosophy annual lecture, given in London on February 18, 1998

published in *Philosophy* vol. 73 no. 285, July 1998, pp. 337-352

Copyright 1998 Cambridge University Press

Thomas Nagel

I

Intuitions based on the first-person perspective can easily mislead us about what is and is not conceivable.¹ This point is usually made in support of familiar reductionist positions on the mind-body problem, but I believe it can be detached from that approach. It seems to me that the powerful appearance of contingency in the relation between the functioning of the physical organism and the conscious mind – an appearance that depends directly or indirectly on the first-person perspective – must be an illusion. But the denial of this contingency should not take the form of a reductionist account of consciousness of the usual type, whereby the logical gap between the mental and the physical is closed by conceptual analysis – in effect, by analyzing the mental in terms of the physical (however elaborately this is done – and I count functionalism as such a theory, along with the topic-neutral causal role analyses of mental concepts from which it descends).

In other words, I believe that there is a necessary connection in both directions between the physical and the mental, but that it cannot be discovered a priori. Opinion is strongly divided on the credibility of some kind of functionalist reductionism, and I won't go through my reasons for being on the antireductionist side of that debate. Despite significant attempts by a number of philosophers to describe the functional manifestations of conscious mental states, I continue to believe that no purely functionalist characterization of a system entails – simply in virtue of our mental concepts – that the system is conscious.

So I want to propose an alternative. In our present situation, when no one has a plausible answer to the mind-body problem, all we can really do is to try to develop various alternatives

¹ See Sydney Shoemaker (1996), *The First-Person Perspective*, in *The First-Person Perspective and Other Essays*, Cambridge University Press.

one of which may prove in the long run to be an ancestor of a credible solution. This is a plea for the project of searching for a solution that takes conscious points of view as logically irreducible to, but nevertheless necessarily connected with, the physical properties of the organisms whose points of view they are. Consciousness should be recognized as a conceptually irreducible aspect of reality that is necessarily connected with other equally irreducible aspects – as electromagnetic fields are irreducible to but necessarily connected with the behavior of charged particles and gravitational fields with the behavior of masses, and vice versa. But the task of conceiving how a necessary connection might hold between the subjective and the physical cannot be accomplished by applying analogies from within physical science. This is a new ballgame. Yet I believe it is not irrational to hope that some day, long after we are all dead, people will be able to observe the operation of the brain and say, with true understanding, “That's what the experience of tasting chocolate looks like from the outside.”

Of course we already know what it looks like from far enough outside: the subject taking the first reverent mouthful of a hot fudge sundae, closing his eyes in rapture, and saying “Yum.” But I have in mind some view or representation of the squishy brain itself, which in light of our understanding we will be able to *see* as tasting chocolate. While that is at the moment inconceivable, I think that it is what we would have to have to grasp what must be the truth about these matters. My reading of the situation is that our inability to come up with an intelligible conception of the relation between mind and body is a sign of the inadequacy of our present concepts, and that some development is needed. At this point, however, all one can hope to do is to state some of the conditions that more adequate concepts would have to satisfy. One can't expect actually to come up with them.² But I shall begin by describing the present impasse.

II

When we try to reason about the possible relations between things, we have to rely on our conceptual grasp of them. The more adequate the grasp, the more reliable our reasoning will be. Sometimes a familiar concept clearly allows for the possibility that what it designates should also have features not implied by the concept itself – often features very different in kind from those

² My position is fairly close to that of Colin McGinn, but without his pessimism. See for example his essay *Consciousness and the Natural Order*, in *The Problem of Consciousness*, Blackwell, 1991. What I have to say here is also a development of a suggestion in *The View From Nowhere*, Oxford University Press, 1986, pp. 51-53.

directly implied by the concept. Thus ordinary prescientific concepts of kinds of substances, such as water or gold or blood, are in themselves silent with regard to the microscopic composition of those substances but nevertheless open to the scientific discovery, often by very indirect means, of such facts about their true nature. If a concept refers to something that takes up room in the spatiotemporal world, it provides a handle for all kinds of empirical discoveries about the inner constitution of that thing.

On the other hand, sometimes a familiar concept clearly excludes the possibility that what it designates has certain features: for example we do not need a scientific investigation to be certain that the number 379 does not have parents. There are various other things that we can come to know about the number 379 only by mathematical or empirical investigation, such as what its factors are, or whether it is greater than the population of Chugwater, Wyoming, but we know that it does not have parents just by knowing that it is a number. If someone rebuked us for being closed-minded, because we can't predict in advance what future scientific research might turn up about the biological origins of numbers, he would not be offering a serious ground for doubt.

The case of mental processes and the brain is intermediate between these two. Descartes thought it was closer to the second category, and that we could tell just by thinking about it that the human mind was not an extended material thing and that no extended material thing could be a thinking subject. But this is, to put it mildly, not nearly as self-evident as that a number cannot have parents. What does seem true is that the concept of a mind, or of a mental event or process, fails to plainly leave space for the possibility that what it designates should turn out also to be a physical thing or event or process, as the result of closer scientific investigation – in the way that the concept of blood leaves space for discoveries about its composition. The trouble is that mental concepts don't obviously pick out things or processes that take up room in the spatiotemporal world to begin with. If they did, we could just get hold of some of those things and take them apart or look at them under a microscope or subject them to chemical analysis. But there is a prior problem about how those concepts might refer to anything that could be subjected to such investigation: They don't give us the comfortable initial handle on the occupants of the familiar spatiotemporal world that prescientific physical substance concepts do.³

³ See Colin McGinn (1995), *Consciousness and Space*, in *Journal of Consciousness Studies* 2, pp. 220-30

Nevertheless it is overconfident to conclude, from one's inability to imagine how mental phenomena might turn out to have physical properties, that the possibility can be ruled out in advance. We have to ask ourselves whether there is more behind the Cartesian intuition than mere lack of knowledge, resulting in lack of imagination.⁴ Of course it is not enough just to say, "You may be mistaking your own inability to imagine something for its inconceivability." One should be open to the possibility of withdrawing a judgment of inconceivability if offered a reason to think it might be mistaken; but there does have to *be* a reason, or at least some kind of story about how this illusion of inconceivability might have arisen.

If mental events really have physical properties, an explanation is needed of why they seem to offer so little purchase for the attribution of those properties. Still, the kind of incomprehensibility here is completely different from that of numbers having parents. Mental events, unlike numbers, can be roughly located in space and time, and are causally related to physical events, in both directions. The causal facts are strong evidence that mental events have physical properties, if only we could make sense of the idea.⁵

Consider another case where the prescientific concept did not obviously allow for the possibility of physical composition or structure – the case of sound. Before the discovery that sounds are waves in air or another medium, the ordinary concept permitted sounds to be roughly located, and to have properties like loudness, pitch, and duration. The concept of a sound was that of an objective phenomenon that could be heard by different people, or that could exist unheard. But it would have been very obscure what could be meant by ascribing to a sound a precise spatial shape and size, or an internal, perhaps microscopic, physical structure. Someone who proposed that sounds have physical parts, without offering any theory to explain this, would not have said anything understandable. One might say that in advance of the development of a physical theory of sound, the hypothesis that sounds have a physical microstructure would not have a clear meaning.

Nevertheless, at one remove, the possibility of such a development is evidently not excluded by the concept of sound. Sounds were known to have certain physical causes, to be blocked by certain kinds of obstacles, and to be perceptible by hearing. This was already a

⁴ This is the objection that Arnauld made to Descartes, in the fourth set of objections to the *Meditations*.

⁵ Compare Donald Davidson (1980), *Mental Events*, in his *Essays on Actions and Events*, Oxford University Press.

substantial amount of causal information, and it opened the way to the discovery of a physically describable phenomenon that could be identified with sound because it had just those causes and effects – particularly once further features of sound, like variations of loudness and pitch, could also be accounted for in terms of its precise physical character. Yet it is important that *in advance*, the idea that a sound has a physical microstructure would have had no clear meaning. One would not have known how to go about imagining such a thing, any more than one could have imagined a sound having weight. It would have been easy to mistake this lack of clear allowance for the possibility in the concept for a positive exclusion of the possibility by the concept.

The analogy with the case of mental phenomena should be clear. They too occupy causal roles, and it has been one of the strongest arguments for some kind of physicalism that those roles may prove upon investigation to be occupied by organic processes. Yet the problem here is much more serious, for an obvious reason: Identifying sounds with waves in the air does not require that we ascribe phenomenological qualities and subjectivity to anything physical, because those are features of the perception of sound, not of sound itself. By contrast, the identification of mental events with physical events requires the unification of these two types of properties in a single thing, and that remains resistant to understanding. The causal argument for identification may make us believe that it is true, but it doesn't help us to understand it, and in my view, we really shouldn't believe it unless we can understand it.

The problem lies in the distinctive first-person/third-person character of mental concepts, which is the grammatical manifestation of the subjectivity of mental phenomena. Though not all conscious beings possess language, our attribution of conscious states to languageless creatures implies that those states are of the kind that in the human case we pick out only through these distinctive concepts, concepts which the subject applies in his own case without observation of his body.

They are not pure first-person concepts: To try to detach their first-person application from the third person results in philosophical illusions: e.g. that the subject of my consciousness might have been replaced five minutes ago and all my memories, personality, etc. transferred to a new subject in this same body, without any outwardly or inwardly perceptible sign – without any other physical or psychological change. If the pure first-person idea of 'I' defined an individual,

that would make sense, but it seems reasonably clear that the real idea of 'I' has lost its moorings in this philosophical thought experiment. The point goes back to Kant, who argued that the subjective identity of the consciousness of myself at different times is not sufficient to establish the objective identity of a subject or soul.⁶

That is not to say that I understand just how the first person and the third form two logically inseparable aspects of a single concept – only that they do. This applies to all conscious mental states and events, and their properties. They are subjective, not in the sense that they are the subjects of a purely first-person vocabulary, but in the sense that they can be accurately described only by concepts in which nonobservational first-person and observational third-person attributions are logically inseparable. Such states are modifications of the point of view of an individual subject.

The problem, then, is how something that is an aspect or element of an individual's subjective point of view could also be a physiologically describable event in the brain – the kind of thing which, considered under that description, involves no point of view and no distinctively immediate first-person attribution at all. I believe that as a matter of fact you can't have one without the other, and furthermore that the powerful intuition that it is conceivable that an intact and normally functioning physical human organism could be a completely unconscious zombie is an illusion – due to the limitations of our understanding. Nevertheless those limitations are real. We do not at present possess the conceptual equipment to understand how subjective and physical features could both be essential aspects of a single entity or process.

III

Suppose, as seems likely, that I taste the smoke of my cigar when and only when my brain is in a certain physical state. What gets in the way of the thought that the experiential state of which I am introspectively aware is the physical state? The problem lies in the lack of any conceivable internal connection between a modification of my subjective point of view and a modification of the physico-chemical activity of my brain. The two may correspond extensionally as exactly as you like, but identity requires more than that. If they are the same state, it must be impossible for the one to exist without the other. And while we may have good

⁶ See *Critique of Pure Reason*, A 363-4: the Paralogisms of Pure Reason.

empirical reasons to believe that that is true, the *understanding* of such an impossibility requires that the necessity of the connection between the two become intellectually transparent to us. In the case of conscious states and physiological states, it isn't just that we don't see such a necessary connection: it seems in advance that a necessary connection between two such different things is unimaginable. They seem logically unrelated.

It is very different from trying to imagine the possibility of a physico-chemical analysis of embryonic development, before one has the slightest inkling of what the analysis might be. In that case we can nevertheless understand the possibility in the abstract: there are no grounds for ruling it out a priori, since we know in general terms what it is to exhaustively analyze a large-scale process into its microscopic or submicroscopic components. But that model will not help us with the mind and the brain. We are entitled to identify a familiar physical phenomenon like fire with a theoretically described process at the atomic level because the macroscopic properties of fire, including its causes and effects, can be derived from the microscopic account – can be seen to be necessitated by it. The chemical analysis of fire and air reveals, for example, that it is a necessary truth that a fire confined to a small, airtight space will quickly go out. But we cannot see how a detailed account of what is going on in the brain could exhaustively explain the taste of a cigar – not even if we could see how it explained all the physical effects of such an experience. So long as this explanatory gap remains, the identification of the states remains problematic.⁷

This does seem to call for some revision in our way of conceiving of mind, or matter, or both. The difficulty is to do this without denying what is in front of your nose. What we need is not a reductionist or eliminative revision but an expansionist one. By this I mean a conception that will permit subjective points of view to have an objective physical character *in themselves*. The reason such an expansion does not seem to me out of the question is that it doesn't involve a contradiction with the essential nature of subjective experience. Our problem is that there is no room for a necessary connection with physiology in the space of possible development defined by the concept of mind. But that does not rule out the possibility of a successor concept of mind which will both preserve the essential features of the original and be open to the discovery of

⁷ Joseph Levine has stressed this point, though I believe he is too ready to describe the problem as epistemological rather than metaphysical. See *Materialism and Qualia: The Explanatory Gap*, in *Pacific Philosophical Quarterly*, 1983.

such connections. This kind of thing happens all the time in the course of scientific history, as with the concepts of sound, or element, or species, or space, or number. A new concept is developed to talk about the same things as the old, one that includes most of the features of the old concept but puts them in a relation to one another and to other features that is new, and that makes it possible to see or explore further connections.

Without such an expanded conception of the mental there is no prospect of overcoming the explanatory gap. But it seems to me entirely possible that a more adequate conception of subjective experience should see it as the phenomenological 'inside' of certain physical processes, and that what we need is to dismantle the barriers to such a conception. If it became possible for us to think of experiences as essentially what it is like subjectively for a physical organism to be in some physiological state, the way would be open for the discovery of a posteriori necessary truths about what physiological state a particular kind of experience is. But the big step is the first one, of expanding the concept of experience by the recognition that what it contains explicitly – including its behavioral or functional implications – gives an incomplete account of the nature of experience.

The ordinary concept of water is unsaturated, so to speak, since it contains a blank space to be filled in by the discovery of the real, and essential, chemical composition of water. Just as we make room for the possibility of such discovery by denying that the manifest properties of water exhaust its nature, so we can open the possibility of an a posteriori answer to the mind-body problem by denying that the manifest properties of experience exhaust its nature. This means thinking of experiences, contrary to intuition, as events whose full nature is not revealed to experience – and more generally, thinking of the mind, contrary to Cartesian intuition, as only partially available, even in principle, to introspection. If we can do this without denying the phenomenology or reducing it to something else, we will be on the first step toward an expansionist but still nondualist response to the mind-body problem. This is so far pure fantasy, but it is the fantasy of a theoretical identification of mental events with an inner constitution that includes but is not exhausted by their introspectible or manifest character.

The search for the possible form of a theory of the relation between mind and brain has to continue, and if there can be no such theory, that too requires explanation. I believe that the explanatory gap in its present form cannot be closed – that so long as we work with our present

mental and physical concepts no transparently necessary connection will ever be revealed, between physically described brain processes and sensory experience, of the logical type familiar from the explanation of other natural processes by analysis into their physico-chemical constituents. We have good grounds for believing that the mental supervenes on the physical – i.e. that there is no mental difference without a physical difference. But pure, unexplained supervenience is not a solution but a sign that there is something fundamental we don't know. We cannot regard pure supervenience as the end of the story because that would require the physical to necessitate the mental without there being any answer to the question *how* it does so. But there *must* be a 'how,' and our task is to understand it. An obviously systematic connection that remains unintelligible to us calls out for a theory.

IV

The subjectivity of consciousness seems to block all reductionist proposals because, given any physicalist or functionalist description, however sophisticated, it *seems* logically possible that there should be an organism or system satisfying those conditions but nevertheless lacking any subjective point of view – a *zombie*, in current jargon. That is, there seems nothing to bar us from conceiving positively the existence of any kind of physical or functionally organized system, and then conceiving negatively that there is no subjective experience connected with it.⁸ The two types of conception are so completely unrelated that the first seems incapable of ruling out the second: All we have to do is imagine the physical system from the outside, and then imagine it from the inside – as not having any inside in the experiential sense.

I said at the outset that we should be very cautious in the use of intuitions that depend on the first-person point of view, and that includes intuitions about persons or beings other than ourselves that depend on taking up their point of view in imagination. In this case the very disparity between the two forms of conception that gives rise to the strong intuition of conceivability should make us suspicious. The absence of any conceptual connection when phenomena are grasped by such disparate concepts may conceal a deeper necessary connection that is not yet conceptual because not accessible to us by means of our present forms of thought.

I still think we can rely on such thought experiments to refute the most common types of

⁸ This argument has recently been given much prominence by David Chalmers; see *The Conscious Mind*, Oxford University Press, 1996.

conceptual reductionism, however. Even if there is some kind of entailment of the mental by the physical-functional, it is not analytic or definitional: There is no hidden verbal contradiction in the description of a zombie – even if in reality a zombie is logically impossible. There remains the question of a possible weaker conceptual link in the other direction – from the mental to the physical – through the necessity of public criteria for the application of mental concepts, which go with their distinctive first-person/third-person character. But while such public criteria may be necessary for the operation of mental concepts, they are not sufficient, and their logical relation to the reference of those concepts is not obvious. In any case, the public criteria are functional rather than physiological, on the usual interpretation, and what I am talking about here is the relation between mental states and the brain, not between mental states and behavior. Let me try to say what is wrong with conceptual arguments against the necessity of the former relation. I shall return to the private language issue later.

The following things seem *prima facie* conceivable which are pretty certainly impossible in a very strong sense, namely:

- (1) a living, behaving, physiologically and functionally perfect human organism that is nevertheless completely lacking in consciousness, i.e. a zombie;
- (2) a conscious subject with an inner life just like ours that behaves and looks just like a human being but has electronic circuitry instead of brains.

To repeat, I believe the apparent conceivability of these things reveals something about our present concepts but not about what is really possible. Analytic psychophysical reductionism is false, but there is independent reason to believe that these are not true logical possibilities, and if so, our concepts are missing something. Our current concepts don't lead to contradiction – it's not as bad as that – but they fail to reveal a logical impossibility.

Contrast these thought experiments with the a priori inconceivability of a number having parents. The latter involves a straightforward clash between concepts, not merely a disparity. No number could enter into the kind of biological relation with a predecessor that is a necessary condition of being a child or offspring. In that case we see a contradiction between the conditions of numberhood and the conditions of being the child of anything or anyone. In the relation of consciousness to the physical world, by contrast, our concepts fail to reveal a necessary connection, and we are tempted to conclude to the absence of any such connection.

Our intuition is of a logical compatibility, not a logical incompatibility. We conceive the body from outside and the mind from inside, and see no internal connection, only an external one of correlation or perhaps causation. But in spite of the vividness of the intuition, I believe that it reflects our conceptual limitations rather than the truth: The difference between the modes of conception is so great that there is every reason to suspect that we would be unable to see an internal necessary connection even if there were one.

Conceivability and inconceivability are the main evidence we have for possibility and necessity, but they can be misleading, and conceivability that depends on the relation between first and third person reference is particularly treacherous terrain. It may be that the physical description of the brain states associated with consciousness is an incomplete account of their essence – that it is merely the outside view of what we recognize from within as conscious experience. If anything like that is true, then our present conceptions of mind and body are radically inadequate to the reality, and do not provide us with adequate tools for a priori reasoning about them. That poses the general question of how we can attempt to develop conceptions that reflect the actual necessary connections and are therefore reliable tools for reasoning, and what determines whether there is hope of developing such concepts for a domain where we do not yet have them. After all, humans did not always have logical, geometrical, and arithmetical concepts, but had to develop them.

V

Obviously we cannot will a new conceptual framework into existence. It has to result from trying to think, in light of the evidence, about the subject we want to understand, and devising concepts that do better justice to it than the ones we have. Considered as a form of revisionism rather than analysis, the physicalist-functionalist movement in philosophy of mind might be thought to have had a similar aim, but I believe it has failed because it is too conservative: It has tried to reinterpret mental concepts so as to make them tractable parts of the framework of physical science. What is needed instead is a search for something more unfamiliar, something which starts from the conceptual unintelligibility, in its present form, of the subjective-objective link. The enterprise is one of imagining possibilities: Identity theorists like Smart, Armstrong, and Lewis tried to explain how the identity of mental with physical states

could be a contingent truth; I am interested in how some sort of mind-brain identity might be a necessary truth. That would require not only the imagination of concepts that might capture the connection, but also some account of how our existing concepts would have to be related to these and to one another. Specifically, we must think about the first-person and third-person conditions of application of mental concepts, and the physical concepts used to describe the brain. What would something be like to which *all* of these applied noncontingently?

If we think of a standard case of reference-fixing (in Kripke's sense) like 'water,' we observe that, prescientifically, the reference is fixed not only by contingent features of the substance, such as its being what falls from the sky as rain and fills the lakes and oceans, what comes out of the tap, and what we bathe in – but also by noncontingent manifest features, such as its density, its liquidity at room temperature and its propensity to freeze or boil under certain conditions. We can adopt a reference-fixing model (rather than the reductionist model) for the relation between functional states and mental states provided we do not think of the functional states as *contingent* reference-fixers. I think the right view is not that functional-role characterizations contingently pick out inner states whose intrinsic nature must be physically or phenomenologically specified, but rather that *all* the types of features by which we standardly identify mental states, from within and from without, are noncontingent features of those states, and that their physiological nature is *also* noncontingent.

That could be true only if they were identified with neurophysiological states whose connection, at least dispositionally, with characteristic functional roles in the organism was noncontingent. This does not seem out of the question: Integrated visual and motor systems in the brain might be part of the specification of the neurophysiological nature of color phenomenology, for example. What I am saying is not incompatible with Kripke's claim that the reference of a term like 'pain' is fixed by an essential property of the referent, namely its immediate phenomenological quality – but it goes beyond this. My suggestion is that the immediate phenomenological quality, the functional role, and the physiological basis of pain are all essential properties of it, and that the apparent conceivability of their separation is an illusion.

To be precise, I conjecture that being a pain entails (nonanalytically) all three features – functional, phenomenological, and physiological – but that only the latter two entail pain. This is because the phenomenological and the physical entail the functional (and one another) whereas

the functional does not entail them. It is parallel to the case of water: There could be a watery liquid that wasn't the compound H₂O and therefore wasn't water; but being H₂O entails having the essential gross properties of water. Similarly, it is possible that there could be a state functionally equivalent to pain in a mechanism with a completely different internal constitution, but if it were both physically and phenomenologically different, it would not be the same mental state and would not be pain. Phenomenological and physiological features of mental states entail their functional features, but not vice versa.

So the proposal is that mental states may have a tripartite essence – phenomenological, functional, and physiological – but we still don't understand how this could be, since our modal intuitions go against it. In particular, we still have to deal with the apparent conceivability of an exact physical-functional replica of a conscious human being that nevertheless has no phenomenological 'interior' at all – a zombie. It is clear that first-person imagination is doing the work in this thought experiment. The task of defending a necessary connection between the physical and the phenomenological would require some account of how a connection that is in fact internal remains stubbornly external from the point of view of our understanding. But I believe that can be done.

VI

Suppose I think about the taste of the cigar I am now smoking. What I must do first is to regard the experience as a state of myself of whose subjective qualities I am immediately aware, and which has certain publicly observable functional relations to stimuli and discriminatory capacities. There is already a natural illusion of contingency with respect to these latter relations, because they are concealed in my introspective identification of the experience. But this introspective identification is itself one of those mental acts that – because of the first-person/third-person link – cannot be logically separated from its functional connections (for example the capacity to distinguish this taste from that of a cigarette). Recognizing this, I can see that the Cartesian thought-experiment of imagining myself having this experience without having a body at all is an unreliable guide to what is really possible. It depends on the concealment of the necessary conditions of reference of the phenomenological concept that I am employing to think about the experience. (This is the anti-private-language point.)

But now what of the relation between the experience and its physiological basis? Here I seem to be able to imagine either myself or someone else tasting exactly this flavor of cigar – and its having all the usual functional connections as well – although my brain or the other person's brain is in a completely different state. Indeed it seems imaginable, though unlikely, that when I offer a friendly cigar to an exotic visitor from outer space who has a completely different physiology, it should taste the same to him. But here too the imagination is a poor guide to possibility, because it relies on an assumption of the completeness of the manifest conditions of reference of the concept (now taken to include functional conditions).

The first thing to acknowledge is that if there were a necessary connection between the phenomenology and the physiology of tasting a cigar, it would not be evident a priori on the basis of the ordinary concept of that experience, since the possession of that concept involves no awareness of anything about the brain. It isn't just that, like the behavioral connections, the relation to the brain is *hidden from view* in my first-person use of the concept: The relation is completely absent from the concept, and cannot be retrieved by philosophical analysis. Nevertheless, if there is such a relation, having the full concept (including the first person aspect) would require having a brain, indeed a brain with exactly the right physiological characteristics, and the brain would be directly involved in the act of imagination – though its involvement would be completely outside the range of my awareness in employing the concept. To imagine a mental state from the inside would be what I have called an act of *sympathetic* imagination – putting myself in a conscious state *resembling* the thing imagined – and it would be impossible to do this without putting my brain in a corresponding physical state.⁹

This shows that I cannot rely on the apparent imaginability of the separation of phenomenology and physiology to establish the contingency of the relation, since I can know in advance that this act of imagination would seem subjectively the same whether the relation were contingent or necessary. If the relation is necessary, then I have not really succeeded in imagining the phenomenology without the physiology. The imagination here is essentially ostensive, and I cannot point to one without pointing to the other. But now, how am I to form the conception that the relation might actually be necessary – as opposed to merely acknowledging

⁹ See “What Is It Like to Be a Bat?” (1974), in *Philosophical Review*, fn. 11. This was an earlier response to the modal argument against materialism. See also Christopher Hill, “Imaginability, Conceivability, Possibility and the Mind-Body Problem,” *Philosophical Studies*, 1995.

that I can't discover a priori that it isn't? I have to think that these two ways of referring – by the phenomenological concept and the physiological concept – pick out a single referent, in each case rigidly, but that the logical link cannot be discovered by inspecting the concepts directly: rather it goes only through their common link to the referent itself.

The idea would have to be, then, that there is a single event to which I can refer in two ways, both of them via concepts that apply to it noncontingently. One is the mental concept that I am able to acquire in both first and third person applications because I am a subject of this state, which has the special character of consciousness and introspective accessibility – the state of tasting a cigar. The other is a (so far unspecified) physiological concept that describes the relevant physical state of the brain. To admit the possibility of a necessary connection here, we have to recognize that the mental concept as it now operates has nothing to say about the physiological conditions for its own operation, and then consider the hypothesis of a successor concept that leaves a place for such a condition – a place that can be filled only a posteriori, by a theory of the third type of event that admits these two types of access, internal and external.

VII

What will be the point of view, so to speak, of such a theory? If we could arrive at it, it would render transparent the relation between mental and physical, not directly, but through the transparency of their common relation to something that is not merely either of them. Neither the mental nor the physical point of view will do for this purpose. The mental will not do because it simply leaves out the physiology, and has no room for it. The physical will not do because while it includes the behavioral and functional manifestations of the mental, this doesn't, in view of the falsity of conceptual reductionism, enable it to reach to the mental concepts themselves. The right point of view would be one which, contrary to present conceptual possibilities, included both subjectivity and spatiotemporal structure from the outset, all its descriptions implying both these things at once, so that it would describe inner states and their functional relations to behavior and to one another from the phenomenological inside and the physiological outside simultaneously – not in parallel. The mental and physiological concepts and their reference to this same inner phenomenon would then be seen as secondary and each partial in its grasp of the phenomenon: Each would be seen as referring to something that extends beyond its grounds of

application.

The difficulty is that such a viewpoint cannot be constructed by the mere conjunction of the mental and the physical. It has to be something genuinely new, otherwise it will not possess the necessary unity. Truly necessary connections could be revealed only by a new theoretical construction, realist in intention, contextually defined as part of a theory that explained both the familiarly observable phenomenological and the physiological characteristics of these inner events. Its character would have to be inferred from what it was introduced to explain – like the electromagnetic field, gravity, the atomic nucleus, or any other theoretical postulate. This could only be done in the context of a theory containing real laws and not just dispositional definitions, otherwise the theoretical entity would not have independent reality.

To summarize. The conjecture is essentially this: that even though no transparent *and direct* explanatory connection is possible between the physiological and the phenomenological, but only an empirically established extensional correlation, we may hope and ought to try as part of a scientific theory of mind to form a third conception that does directly entail both the mental and the physical, and through which their actual necessary connection with one another can therefore become transparent to us. Such a conception will have to be created; we won't just find it lying around. All the great reductive successes in the history of science have depended on theoretical concepts, not natural ones – concepts whose whole justification is that they permit us to replace brute correlations with reductive explanations. At present such a solution to the mind-body problem is literally unimaginable, but it may not be impossible.