

# The Mind–Body Problem: An Overview

*Kirk Ludwig*

*I have said that the soul is not more than the body,  
And I have said that the body is not more than the soul,  
And nothing, not God, is greater to one than one's self is.*  
Walt Whitman

## 1.1 Introduction

Understanding the place of thought and feeling in the natural world is central to that general comprehension of nature, as well as that special self-understanding, which are the primary goals of science and philosophy. The general form of the project, which has exercised scientists and philosophers since the ancient world, is given by the question, ‘What is the relation, in general, between mental and physical phenomena?’ There is no settled agreement on the correct answer. This is the single most important gap in our understanding of the natural world. The trouble is that the question presents us with a problem: each *possible* answer to it has consequences that appear unacceptable. This problem has traditionally gone under the heading ‘The Mind–Body Problem.’<sup>1</sup> My primary aim in this chapter is to explain in what this traditional mind–body problem consists, what its possible solutions are, and what obstacles lie in the way of a resolution.

The discussion will develop in two phases. The first phase, sections 1.2–1.4, will be concerned to get clearer about the import of our initial question as a precondition of developing an account of possible responses to it. The second phase, sections 1.5–1.6, explains how a problem arises in our attempts to answer the question we have characterized, and surveys the various solutions that can be and have been offered.

More specifically, sections 1.2–1.4 are concerned with how to understand the basic elements of our initial question – how we should identify the mental, on the

one hand, and the physical, on the other – and with what sorts of relations between them we are concerned. Section 1.2 identifies and explains the two traditional marks of the mental, consciousness and intentionality, and discusses how they are related. Section 1.3 gives an account of how we should understand ‘physical’ in our initial question so as not to foreclose any of the traditional positions on the mind–body problem. Section 1.4 then addresses the third element in our initial question, mapping out the basic sorts of relations that may hold between mental and physical phenomena, and identifying some for special attention.

Sections 1.5–1.6 are concerned with explaining the source of the difficulty in answering our initial question, and the kinds of solutions that have been offered to it. Section 1.5 explains why our initial question gives rise to a problem, and gives a precise form to the mind–body problem, which is presented as a set of four propositions, each of which, when presented independently, seems compelling, but which are jointly inconsistent. Section 1.6 classifies responses to the mind–body problem on the basis of which of the propositions in our inconsistent set they reject, and provides a brief overview of the main varieties in each category, together with some of the difficulties that arise for each. Section 1.7 is a brief conclusion about the source of our difficulties in understanding the place of mind in the natural world.<sup>2</sup>

## 1.2 Marks of the Mental

The suggestion that consciousness is a mark of the mental traces back at least to Descartes.<sup>3</sup> Consciousness is the most salient feature of our mental lives. As William James put it, “The first and foremost concrete fact which every one will affirm to belong to his inner experience is the fact that *consciousness of some sort goes on*” (James 1910: 71). A state or event (a change of state of an object<sup>4</sup>) is mental, on this view, if it is *conscious*. States, in turn, are individuated by the properties the having of which by objects constitutes their being in them.

Identifying consciousness as a mark of the mental only pushes our question one step back. We must now say what it is for something to be conscious. This is not easy to do. There are two immediate difficulties. First, in G. E. Moore’s words, “the moment we try to fix our attention upon consciousness and to see *what*, distinctly, it is, it seems to vanish: it seems as if we had before us a mere emptiness . . . as if it were diaphanous” (1903: 25). Second, it is not clear that consciousness, even if we get a fix on it, is understandable in other terms. To say something substantive about it is to say something contentious as well. For present purposes, however, it will be enough to indicate what we are interested in in a way that everyone will be able to agree upon. What I say now then is not intended to provide an analysis of consciousness, but rather to draw attention to, and to describe, the phenomenon, in much the same way a naturalist would draw attention to a certain species of insect or plant by pointing one out, or describing

conditions under which it is observed, and describing its features, features which anyone in an appropriate position can himself confirm to be features of it.

First, then, *we* are conscious when we are awake rather than in dreamless sleep, and, in sleep, *when* we dream. When we are conscious, we have conscious states, which we can discriminate, and remember as well as forget. Each conscious mental state is a mode, or way, of being conscious. Knowledge of our conscious mental states, even when connected in perceptual experiences with knowledge of the world, is yet distinct from it, as is shown by the possibility of indistinguishable yet non-veridical perceptual experiences. Conscious mental states include paradigmatically perceptual experiences, somatic sensations, proprioception, pains and itches, feeling sad or angry, or hunger or thirst, and occurrent thoughts and desires. In Thomas Nagel’s evocative phrase, an organism has conscious mental states if and only if “there is something it is like to be that organism” (1979b: 166). There is, in contrast, nothing it is like in the relevant sense, it is usually thought, to be a toenail, or a chair, or a blade of grass.

In trying to capture the kinds of discrimination we make between modes of consciousness (or ways of being conscious), it is said that conscious states have a phenomenal or qualitative character; the phenomenal qualities of conscious mental states are often called ‘qualia’. Sometimes qualia are reified and treated as if they were objects of awareness in the way tables and chairs are objects of perception. But this is a mistake. When one is aware of one’s own conscious mental states or their phenomenal qualities, the only object in question is oneself: what one is aware of is a particular modification of that object, a way it is conscious. Similarly, when we see a red apple, we see just the apple, and not the redness as another thing alongside it: rather, we represent the apple we see *as* red.

A striking feature of our conscious mental states is that we have non-inferential knowledge of them. When we are conscious, we know that we are, and we know *how* we are conscious, that is, our modes of consciousness, but we do not infer, when we are conscious, that we are, or how we are, from anything of which we are more directly aware, or know independently.<sup>5</sup> It is notoriously difficult to say what this kind of non-inferential knowledge comes to. It is difficult to see how to separate it from what we think of as the qualitative character of conscious mental states.<sup>6</sup> Arguably this “first-person” knowledge is *sui generis*. There is a related *asymmetry* in our relation to our own and others’ conscious mental states. We do not have to *infer* that we are conscious, but others must do so, typically from our behavior, and cannot know non-inferentially. Others have, at best, “third-person” knowledge of our mental states. These special features of conscious states are connected with some of the puzzles that arise from the attempt to answer our opening question. Consciousness has often been seen as the central mystery in the mind–body problem, and the primary obstacle to an adequate physicalist understanding of the mental.<sup>7</sup>

The other traditional mark of the mental, first articulated clearly by Franz Brentano (1955 [1874], bk 2, ch. 1), is called ‘intentionality’.<sup>8</sup> The adjectival form is ‘intentional’. But this is a technical term, and does not just involve those

states that in English are called ‘intentions’ (such as my intention to have another cup of coffee). Intentionality, rather, is the feature of a state or event that makes it *about* or *directed at* something. The best way to make this clearer is to give some examples. Unlike the chair that I am sitting in as I write, I have various beliefs *about* myself, my surroundings, and my past and future. I believe that I will have another cup of coffee before the day is out. My chair has no corresponding belief, nor any other. Beliefs are paradigmatically intentional states. They represent the world as being a certain way. They can be true or false. This is their particular form of *satisfaction* condition. In John Searle’s apt phrase, they have mind-to-world direction of fit (1983: ch. 1). They are supposed to fit the world. Any state with mind-to-world direction of fit, any representational state, or attitude, is an intentional state (in the technical sense). False beliefs are just as much intentional states as true ones, even if there is nothing in the world for them to be about of the sort they represent. I can think *about* unicorns, though there are none. The representation can exist without what it represents. It is this sense of ‘aboutness’ or ‘directedness’ that is at issue in thinking about intentionality.

There are intentional states with mind-to-world direction of fit in addition to beliefs, such as expectations, suppositions, convictions, opinions, doubts, and so on. Not all intentional states have mind-to-world direction of fit, however. Another important class is exemplified by desires or wants. I believe I will, but also want to have another cup of coffee soon. This desire is also directed at or about the world, and even more obviously than in the case of belief, there need not be anything in the world corresponding. But in contrast to belief, its aim is not to get its content (that I have another cup of coffee soon) to match the world, but *to get the world to match its content*. It has world-to-mind direction of fit. A desire may be satisfied or fail to be satisfied, just as a belief can be true or false. This is its particular form of satisfaction condition. Any state with world-to-mind direction of fit is likewise an intentional state.

Clearly there can be something in common between beliefs and desires. I believe that *I will have another cup of coffee soon*, and I desire that *I will have another cup of coffee soon*. These have in common their content, and it is in virtue of their content that each is an intentional state. (Elements in common between contents, which would be expressed using a general term, are typically called ‘concepts’; thus, the concept of *coffee* is said to be a constituent of the content of the belief *that coffee is a beverage* and of the belief *that coffee contains caffeine*.) The content in each matches or fails to match the world. The difference between beliefs and desires lies in their role in our mental economy: whether their purpose is to change so that their content matches the world (beliefs) or to get the world to change to match their content (desires). States like these with contents that we can express using sentences are called ‘propositional attitudes’ (a term introduced by Bertrand Russell, after the supposed objects of the attitudes, propositions, named or denoted by phrases of the form ‘that *p*’, where ‘*p*’ is replaced by a sentence). Propositional attitudes are individuated by their psychological mode (belief, supposition, doubt, desire, aspiration, etc.) and content. States with world-to-mind direction of fit are

*pro* or, if negative, *con attitudes*. There are many varieties besides desires and wants, such as hopes, fears, likes, dislikes, and so on.

It is not clear that all representational content is fully propositional. Our perceptual experiences, e.g., our visual, auditory, and tactile experiences, represent our environments as being a certain way. They can be *veridical* (correctly represent) or *non-veridical* (incorrectly represent), as beliefs can be true or false. They have mind-to-world direction of fit, hence, representational contents, and intentionality. But it is not clear that all that they represent could be captured propositionally. Attitudes and perceptual experiences might be said to be different currencies for which there is no precise standard of exchange.

Can there be states directed at or about something which do not have full contents? Someone could have a fear of spiders without having any desires directed at particular spiders, though the fear is in a sense directed at or about spiders. Yet a fear of spiders does entail a desire to avoid contact with, or proximity to, spiders: and it is this together with a particular emotional aura which thinking of or perceiving spiders evokes which we think of as the fear of spiders. In any case, we will call this class of states intentional states as well, though their intentionality seems to be grounded in the intentionality of representational, or *pro* or *con* attitudes, which underlie them, or, as we can say, on which they depend.

We may, then, say that an intentional state is a state with a content (in the sense we've characterized) or which depends (in the sense just indicated) on such a state.<sup>9</sup>

A state then is a mental state (or event) if and only if (iff) it is either a conscious or an intentional state (or event). An object is a thinking thing iff it has mental states.

What is the relation between conscious states and intentional states? If the two sorts are independent, then our initial question breaks down into two subquestions, one about the relation of consciousness, and one about that of intentionality, to the physical. If the two sorts are not independent of one another, any answer to the general question must tackle both subquestions at once.

Some intentional states are clearly not conscious states. Your belief that Australia lies in the Antipodes was not a conscious belief (or an *occurrent* belief) just a moment ago. You were not *thinking* that, though you believed it. It was a *dispositional*, as opposed to an *occurrent*, belief. The distinction generalizes to all attitude types. A desire can be *occurrent*, my present desire for a cup of coffee, for example, or *dispositional*, my desire to buy a certain book when I am not thinking about it.<sup>10</sup> This does not, however, settle the question whether intentional and conscious mental states are independent. It may be a necessary condition on our conceiving of dispositional mental states as intentional attitudes that among their manifestation properties are *occurrent* attitudes with the same mode and content. In this case, the strategy of *divide and conquer* will be unavailable: we will not be able to separate the projects of understanding the intentional and the conscious, and proceed to tackle each independently.<sup>11</sup>

Some conscious mental states seem to lack intentionality, for example, certain episodes of euphoria or anxiety. Though typically caused by our beliefs and

desires, it is not clear that they are themselves about anything. Likewise, somatic sensations such as itches and pains seem to have non-representational elements. Typically somatic sensations represent something's occurring in one's body. A headache is represented as in the head, a toe ache as in the toe. But the quality of pain itself, though it be taken to be a biological indicator of, say, damage to the body, in the way that smoke indicates combustion, seems not to have any associated representational content. Pain does not *represent* (as opposed to indicate) damage. And, though we usually wish pain we experience to cease, the desire that one's pain cease, which has representational content, is not the pain itself, any more than a desire for a larger house is itself a house.<sup>12</sup>

### 1.3 The Physical

Characterizing physical phenomena in a way that captures the intention of our initial question is not as easy as it may appear. We cannot say that physical phenomena consist in what our current physics talks about. Physical theory changes constantly; current physical theory may undergo radical revision, as past physical theory has. The mind-body problem doesn't change with passing physical theory. There are at least three other options.

The first is to characterize physical phenomena as what the ultimately correct physical theory talks about, where we think of physical theory as the theory that tells us about the basic constituents of things and their properties. The second is to treat physical phenomena as by definition non-mental. There are reasons to think that neither of these captures the sense of our initial question.

One response to the mind-body problem is that the basic constituents of things have irreducible mental properties. On the first interpretation, such a position would be classified as a version of physicalism (we will give a precise characterization of this at the end of section 1.4), since it holds that mental properties are, in the relevant sense, physical properties. But this position, that the basic constituents of things have irreducible mental properties, is usually thought to be incompatible with physicalism.

The second interpretation in its turn does not leave open the option of seeing mental phenomena as conceptually reducible to physical phenomena. If the physical is non-mental per se, then showing that mental properties are really properties that fall in category *F* would just show that a subcategory of properties in category *F* were not physical properties. But we want the terms in which our initial question is stated to leave it open whether mental properties are conceptually reducible to physical properties. (We will return to what this could come to below.)

A third option is to take physical phenomena to be of a general type exemplified by our current physics. Here we would aim to characterize a class of properties that subsumes those appealed to by past and current physical theories, from the scientific revolution to the present, but which is broad enough to cover properties appealed

to in any extension of our current approach to explaining the dynamics of material objects. This interpretation leaves open the options foreclosed by our first two interpretations, and comports well with the development of concerns about the relation of mental to physical phenomena from the early modern period to the present. It is not easy to say how to characterize the intended class of properties. The core conception of them is given by those qualities classed as *primary qualities* in the seventeenth and eighteenth centuries: size, shape, motion, number, solidity, texture, logical constructions of these, and properties characterized essentially in terms of their *effects* on these (mass and charge, e.g., arguably fall in the last category).<sup>13</sup> It is not clear that this is adequate to cover everything we might wish to include. But it is fair to say that, typically, philosophers have in mind this conception of the physical in posing the question we began with, without having a detailed conception of how to delineate the relevant class of properties.<sup>14</sup>

#### 1.4 Mind–Body Relations

The question of the relation between the mental and the physical can be posed equivalently as about mental and physical properties, concepts, or predicates. A property is a feature of an object, such as being round, or being three feet from the earth’s surface. A concept, as we have said, is a common element in different thought contents expressed by a general term. We deploy concepts in thinking about a thing’s properties. So, corresponding to the property of being round is the concept of being round, or of roundness. When I think that this ball is round, and so think of it as having the property of being round, I have a thought that involves the concept of being round. I am said to bring the ball under the concept of roundness. Predicates express concepts, and are used to attribute properties to objects.<sup>15</sup> Thus, ‘is round’ expresses (in English) the concept of roundness, and is used to attribute the property of being round. We may say it picks out that property. For every property there is a unique concept that is about it, and vice versa. More than one predicate can express the same concept, and pick out the same property, but then they must be synonymous.<sup>16</sup> Corresponding to each property category (mental or physical, e.g.) is a category of concepts and predicates. Thus, any question we ask about the relation of mental and physical properties can be recast as about concepts or predicates, and vice versa.

The basic options in thinking about the relation of mental and physical properties can be explained in terms of the following three sentence forms, where ‘is *M*’ represents a mental predicate, and ‘is *P*’ represents a physical predicate (this is generalizable straightforwardly to relational terms).

- [A] For all  $x$ , if  $x$  is  $P$ , then  $x$  is  $M$
- [B] For all  $x$ , if  $x$  is  $M$ , then  $x$  is  $P$
- [C] For all  $x$ ,  $x$  is  $M$  if and only if (iff)  $x$  is  $P$

Though [C] is equivalent to the conjunction of [A] and [B], it will be useful to state it separately. The relation of the mental to the physical is determined by which instances of [A]–[C] are true or false, and on what grounds. One could hold each to be necessarily true or necessarily false, in one of three senses of “necessity”: conceptual, metaphysical (*so-called*), and nomological.

Two notions that figure prominently in discussions of the mind–body problem can be characterized in this framework. The first is that of reduction, and the second that of supervenience. Each can be conceptual, metaphysical, or nomological. I begin with conceptual reduction and supervenience.

Conceptual necessities are truths grounded in the concepts used to express them. This is the strongest sort of necessity. What is conceptually necessary is so in every metaphysically and nomologically possible world, though not vice versa. Knowledge of conceptual truths can be obtained from reflection on the concepts involved, and need not rest on experience (traditionally, knowledge of one’s own conscious mental states is counted as experiential knowledge). They are thus said to be knowable a priori. Knowledge obtained in this way is a priori knowledge. A proposition known on the basis of experience is known a posteriori, or empirically. Knowledge so based is a posteriori or empirical knowledge. Conceptual truths are not refutable by the contents of any experiences. A sentence expressing (in a language *L*) a conceptual truth is *analytically true (in L)*, or, equivalently, *analytic (in L)* (henceforth I omit the relativization). A sentence is analytic iff its truth is entailed by true meaning-statements about its constituents.<sup>17</sup> For example, ‘None of the inhabitants of Dublin resides elsewhere’, or ‘There is no greatest prime number’ would typically be regarded as analytic.<sup>18</sup>

Conceptual reduction of mental to physical properties, or vice versa, is the strongest connection that can obtain between them. (We say equivalently, in this case, that mental concepts/predicates can be analyzed in terms of physical concepts/predicates, or vice versa.) If a mental property is conceptually reducible to a physical property, then two conditions are met: (a) the instance of [C], in which ‘is *M*’ is replaced by a predicate that picks out the mental property, and ‘is *P*’ by a (possibly complex) predicate that picks out the physical property, is conceptually necessary, and (b) the concepts expressed by ‘is *P*’ are conceptually prior to those expressed by ‘is *M*’, which is to say that we have to have the concepts expressed by ‘is *P*’ in order to understand those expressed by ‘is *M*’, but not vice versa (think of the order in which we construct geometrical concepts as an example). The second clause gives content to the idea that we have effected a reduction, for it requires the physical concepts to be more basic than the mental concepts. A conceptual reduction of a mental property to a physical property shows the mental property to be a species of physical property. This amounts to the identification of a mental property with a physical property. Similarly for the reduction of a physical property to a mental property.

One could hold that instances of [C] were conceptually necessary without holding that either the mental or the physical was conceptually reducible to the other. In this case, their necessary correlation would be explained by appeal to another set

of concepts neither physical nor mental, in terms of which each could be understood. For example, it is conceptually necessary that every triangle is a trilateral, but neither of these notions provides a conceptual reduction of the other.

‘Supervenience’ is a term of art used in much current philosophical literature on the mind–body problem. It may be doubted that it is needed in order to discuss the mind–body problem, but given its current widespread use, no contemporary survey of the mind–body problem should omit its mention. A variety of related notions has been expressed using it. Though varying in strength among themselves, they are generally intended to express theses weaker than reductionism, invoking only sufficiency conditions, rather than conditions that are both necessary and sufficient.<sup>19</sup> Supervenience claims are not supposed to provide explanations, but rather to place constraints on the *form* of an explanation of one sort of properties in terms of another. I introduce here a definition of one family of properties supervening on another, which will be useful for formulating a position we will call ‘physicalism’, and which will be useful later in our discussion of a position on the relation of mental to physical properties known as ‘functionalism’. I begin with ‘conceptual supervenience’.

*F*-properties *conceptually supervene* on *G*-properties iff for any *x*, if *x* has a property *f* from *F*, then there is a property *g* from *G*, such that *x* has *g* and it is *conceptually necessary* that if *x* has *g*, then *x* has *f*.<sup>20</sup>

Conceptual reduction of one family of properties to another implies mutual conceptual supervenience. But the supervenience of one family of properties on another does not imply their reducibility to them.

I will characterize ‘physicalism’ as the position according to which, *whatever mental properties objects have, they conceptually supervene on the physical properties objects have, and whatever psychological laws there are, the physical laws entail them*.<sup>21</sup> This allows someone who thinks that nothing has mental properties, and that there are no mental laws, to count as a physicalist, whatever his view about the conceptual relations between mental and physical properties.<sup>22</sup> The definition here is stipulative, though it is intended to track a widespread (though not universal) usage in the philosophical literature on the mind–body problem.<sup>23</sup> The question whether physicalism is true, so understood, marks a fundamental divide in positions on the mind–body problem.

Nomological necessity we can explain in terms of conceptual necessity and the notion of a natural law. A statement that *p* is nomologically necessary iff it is conceptually necessary that if *L*, it is the case that *p*, where “*L*” stands in for a sentence expressing all the laws of nature, whether physical or not (adding “boundary conditions” to “*L*” yields more restrictive notions). I offer only a negative characterization of metaphysical necessity, which has received considerable attention in contemporary discussion of the mind–body problem. I will argue in section 1.6 that no concept corresponds to the expression “metaphysical necessity” in these contexts, despite its widespread use. For now, we can say that metaphysical

necessity is supposed to be of a sort that cannot be discovered a priori, but which is stronger than nomological necessity, and weaker than conceptual necessity. To obtain corresponding notions of *metaphysical* and *nomological* supervenience, we substitute ‘metaphysically’ or ‘nomologically’ for ‘conceptually’ in our characterization above.

Metaphysical and nomological reduction require that biconditionals of the form [C] are metaphysically or nomologically necessary (but nothing stronger), respectively. But reduction is asymmetric. So we must also give a sense to the idea that one side of the biconditional expresses properties that are more basic. In practice, the question is how to make sense of the asymmetry for metaphysical or nomological reduction of the mental to the physical. There is nothing in the case of metaphysical or nomological necessity that corresponds to conceptual priority. It looks as if the best we can do is to ground the desired asymmetry in physical properties being basic in our general explanatory scheme. This is usually understood to mean that the physical constitutes an explanatorily closed system, while the mental does not. This means that every event can be explained by invoking physical antecedents, but not by invoking mental antecedents.

### 1.5 The Mind–Body Problem

A philosophical problem is a knot in our thinking about some fundamental matter that we have difficulty unraveling. Usually, this involves conceptual issues that are particularly difficult to sort through. Because philosophical problems involve foundational issues, how we resolve them has significant import for our understanding of an entire field of inquiry. Often, a philosophical problem can be presented as a set of propositions all of which seem true on an initial survey, or for all of which there are powerful reasons, but which are jointly inconsistent. This is the form in which the problem of freedom of the will and skepticism about the external world present themselves. It is a significant advance if we can put a problem in this way. For the ways in which consistency can be restored to our views determines the logical space of solutions to it. The mind–body problem can be posed in this way. Historical and contemporary positions on the relation of the mental to the physical can then be classified in terms of which of the propositions they choose to reject to restore consistency.

The problem arises from the appeal of the following four theses.

- 1 *Realism*. Some things have mental properties.
- 2 *Conceptual autonomy*. Mental properties are not conceptually reducible to non-mental properties, and, consequently, no non-mental proposition entails any mental proposition.<sup>24</sup>
- 3 *Constituent explanatory sufficiency*. A complete description of a thing in terms of its basic constituents, their non-relational properties,<sup>25</sup> and relations to

one another<sup>26</sup> and to other basic constituents of things, similarly described (the constituent description) entails a complete description of it, i.e., an account of all of a thing's properties follows from its constituent description.

- 4 *Constituent non-mentalism*. The basic constituents of things do not have mental properties as such.<sup>27</sup>

The logical difficulty can now be precisely stated. Theses (2)–(4) entail the negation of (1). For if the correct fundamental physics invokes no mental properties, (4), and every natural phenomenon (i.e., every phenomenon) is deducible from a description of a thing in terms of its basic constituents and their arrangements, (3), then given that no non-mental propositions entail any mental propositions, (2), we can deduce that *there are no things with mental properties*, which is the negation of (1).

The logical difficulty would be easy to resolve were it not for the fact that each of (1)–(4) has a powerful appeal for us.

Thesis (1) seems *obviously* true. We seem to have direct, non-inferential knowledge of our own conscious mental states. We attribute to one another mental states in explaining what we do, and base our predictions on what others will do in part on our beliefs about what attitudes they have and what their conscious states are. Relinquishing (1) seems unimaginable.

Proposition (2) is strongly supported by the *prima facie* intelligibility of a body whose behavior is like that of a thinking being but which has no mental life of the sort we are aware of from our own point of view. We imagine that our mental states cause our behavior. It seems conceivable that such behavior results from other causes. Indeed, it seems conceivable that it be caused from exactly the physical states of our bodies that we have independent reasons to think animate them without the accompanying choir of consciousness. It is likewise supported by the *prima facie* intelligibility of non-material thinking beings (such as God and His angels, whom even atheists have typically taken to be conceivable). Thus, it seems, *prima facie*, that having a material body is neither conceptually necessary nor sufficient for having the sorts of mental lives we do.

Thought experiments ask us to imagine a possibly contrary to fact situation and ask ourselves whether it appears barely to make sense (not just whether it is compatible with natural law) that a certain state of affairs could then obtain. We typically test conceptual connections in this way. For example, we can ask ourselves whether we can conceive of an object that is red but not extended. The answer is 'no'. We can likewise ask whether we can conceive of an object that is red and shaped like a penguin. The answer is 'yes'. This provides evidence that the first is conceptually impossible – ruled out by the concepts involved in its description – and that the second is conceptually possible – not ruled out by the concepts involved. No one is likely to dispute the results here.<sup>28</sup> But we can be misled. For example, it may seem easy to conceive of a set that contains all and only sets which do not contain themselves (the Russell set). For it is easy to conceive a set which contains no sets, and a set which contains sets only, and so

it can seem easy to conceive of a special set of sets whose members are just those sets not containing themselves. But it is possible to show that this leads to a contradiction. Call the set of all sets that do not contain themselves ‘ $R$ ’. If  $R$  is a member of  $R$ , it fails to meet the membership condition for  $R$ , and so is not a member of itself. But if it is not a member of itself, then it meets the membership condition and so is a member of itself. So, it is a member of itself iff it is *not*, which is a contradiction, and necessarily false. There cannot be such a set.<sup>29</sup> Thus, something can seem conceivable to us even when it is not. In light of this, it is open for someone to object that despite the apparent intelligibility of the thought experiments that support (2), we have made some mistake in thinking them through.<sup>30</sup>

Proposition (3) is supported by the success of science in explaining the behavior of complex systems in terms of laws governing their constituents. While there are still many things we do not understand about the relation of micro to macro phenomena, it looks as if the techniques so far applied with success can be extended to those features of complex systems we don’t yet understand fully in terms of their constituents’ properties – with the possible exception of psychological phenomena. Proposition (3) expresses a thought that has had a powerful ideological hold on our the scientific worldview, that nature is ultimately intelligible as a kind of vast machine, a complex system a complete understanding of which can be obtained by analyzing its structure and the laws governing the properties of its parts. “It has been,” in E. O. Wilson’s words, “tested in acid baths of experiment and logic and enjoyed repeated vindication” (1998: 5). This thought motivates much scientific research, and to give it up even with respect to a part of the natural world would be to give up a central methodological tenet of our current scientific worldview. It would be to admit that nature contains some basic element of arbitrariness, in the sense that there would be features of objects that were not explicable as arising from their manner of construction.

Finally, proposition (4) is supported also by the success of physics (so far) in accounting for the phenomena that fall in its domain without appeal to any mental properties. In the catalog of properties of particle physics, we find mass, charge, velocity, position, size, spin, and the like, but nothing that bears the least hint of the mental, and nothing of that sort looks to be required to explain the interaction and dynamics of the smallest bits of matter.<sup>31</sup> It can seem difficult even to understand what it would be to attribute mental properties to the smallest constituents of matter, which are incapable of any of the outward signs of mental activity.

This then is the mind–body problem. Propositions (1)–(4) all seem to be true. But they cannot all be, for they are jointly inconsistent. That is why our initial question, “What is the relation, in general, between mental and physical phenomena?” gives rise to a philosophical problem. Each answer we might like to give will involve rejecting one of our propositions (1)–(4); yet, considered independently, each of these propositions seems to be one we have good reasons to accept.

## 1.6 The Logical Space of Solutions

Proposed solutions to the mind–body problem can be classified according to which of (1)–(4) they reject to restore consistency. There are only four basic positions, since we seek a minimal revision. To reject (1) is to adopt *irrealism* or *eliminativism* about the mental. To reject (2) is to adopt *conceptual reductionism* for the mental. This includes neutral monism, psychophysical identity theories, functionalism, and functionalism-cum-externalism. To reject (3) is to adopt *conceptual anti-reductionism*, but not ontological anti-reductionism. Neutral emergentism and emergent materialism fall into this category. To reject (4) is to adopt *ontological anti-reductionism* in addition to conceptual anti-reductionism. This subsumes varieties of what might be called ‘mental particle theories’, and includes substance dualism, idealism, panpsychism, double (or dual) aspect theories (on a certain conception), and what I will call ‘special particle theories’.

We take up each in reverse order, since this represents their historical development. I primarily discuss views on the mind–body problem from the beginning of the modern period to the present, though in fact all the basic positions except eliminativism were anticipated in antiquity.<sup>32</sup>

### 1.6.1 *Ontological anti-reductionism*

Rejecting proposition (4), the non-mental character of the basic constituents of things, has been historically the most popular position. The generic view, according to which some *basic* constituents of things *as such* have mental properties, may be called ‘the mental particle theory’. These may be further divided into *pure* and *mixed* mental particle theories, according to whether the mental particles are thought to have only mental, or to have mental and physical properties, and then, divided again according to whether all or only some things have mental properties (*universal* vs. *restricted*).

The most prominent, and historically important, view of this sort is substance dualism, which traces back to the ancient view of the soul as a simple substance.<sup>33</sup> Substance dualism holds that there are both material substances and mental substances: the former have only physical properties, and none mental, the latter only mental properties, and none physical. This is a restricted pure mental particle theory. Descartes (1985 [1641]) is the most prominent of the early modern defenders of dualism. The appeal of dualism lies in part in its ability to find a place for irreducible mental properties in a world that seems largely to be explainable as a mechanical system reducible to parts which themselves are exhaustively characterized in terms of their primary qualities. Descartes wrote at the beginning of the scientific revolution, and was himself a major proponent of the new ‘mechanical philosophy’, whose fundamental assumptions provide those for modern physics.

Dualism was Descartes's answer to the problem the mechanical philosophy presents for finding a place for mind in the natural world.

Descartes has had such an enormous influence on the development of the western tradition in philosophy that it will be useful to review briefly his official arguments for dualism. This sets the stage for subsequent discussions of the mind-body problem. To explain Descartes's arguments, however, we must first get clearer about the notion of a substance. This notion, central to philosophical discussion in the seventeenth and eighteenth centuries,<sup>34</sup> traces back to Aristotle's characterization of it as "that which is neither said of a subject nor in a subject" (*Categories* (*Cat*) 1b2–5; in 1984: 4). This is the conception of a substance as a property bearer, something that undergoes and persists through change: "A substance . . . numerically one and the same, is able to receive contraries . . . pale at one time and dark at another" (*Cat* 4a19–21; in 1984: 7). This gave rise in medieval philosophy (in scholasticism, the tradition to which the recovery of Aristotle's works gave rise) to the view of substances as independent existents, because of the contrast with properties, which were thought to exist only in a subject, not independently. Descartes gives two characterizations of substance. One is as that which is absolutely independent of everything else. This generalizes the scholastic notion. Descartes held that, on this conception, God is the only substance, since everything depends on God for its existence. But Descartes admits substances as property bearers in a subsidiary sense, and allows two fundamentally different kinds in addition to God: thinking and corporeal substances (*Princ.* 1644, I.51–2; in 1985, vol. I: 210). Henceforth I restrict attention to the latter sort. A central feature of Descartes's theory of substance kinds is that each different substance kind has a principal individuating attribute, of which every other property of a substance of the kind is a modification: *extension*, for corporeal substances, and *thought*, for thinking substances (*Princ.* 1644, I.53–4; in 1985, vol. I: 210–11). This feature of the theory, often overlooked in introductory discussions, is essential for a correct understanding of the force of Descartes's arguments for substance dualism.

The doctrine that each substance has a principal attribute forces the individuating and essential property of a substance kind to be a fundamental way of being something, or a categorical property. A categorical property is a determinable but not a determinate. A determinable is a property an object can have in different ways, and must have in some particular way, as, e.g., being colored. Something can be colored by being blue, or green, or red, and so on, and if colored must be colored in some determinate way (hence the terminology, 'determinable', 'determinate'). Extension and thought Descartes conceived as determinables, and they are not themselves apparently determinates of any other determinable property.<sup>35</sup>

With this theory in place, there is an easy argument to mind-body dualism. If there are two most general ways of being, and things that have them, it follows immediately that there are two kinds of substance. Descartes argued that he had a clear and distinct conception of himself as a thinking thing, a thing that at least can exist independently of his body, and likewise a clear and distinct conception

of a corporeal object as a solely extended thing, a thing that can at least exist without thinking, and, moreover, that these conceptions are complete and not in need of appeal to any more general conception of a kind.<sup>36</sup> From this, it follows that thinking and extension are categorical properties. From the theory of substances, it follows that thinking and extended substances are necessarily distinct.

The argument is unquestionably valid: necessarily, if its premises are true, so is its conclusion. Whether we should accept its premises (and so whether it is sound, i.e., has true premises in addition to being valid) is less clear. Its weakest premise is the assumption that distinct kinds of substance must have only one categorical attribute. It is unclear why Descartes held this. The thought that substances are property bearers provides insufficient support. Even Spinoza, who was heavily influenced by Descartes, objected that precisely because mental and corporeal properties are conceptually independent, there can be no barrier to one substance possessing both attributes (*Ethics* IP10 Scholium; in Spinoza 1994: 90). And, as P. F. Strawson (1958) has observed, we routinely attribute to the very same thing, persons, both material and mental properties: I walk, and sleep, as well as think and feel.

Descartes endorsed causal interactionism between mental and material substance to explain why our limbs move in accordance with what we want to do, and how we are able to correctly perceive things in our bodies' physical surroundings. Some philosophers, including many of Descartes's contemporaries, have objected that we cannot conceive of causal interaction between such fundamentally different kinds of substance as mind and body, the latter in space, the former not. (Though it is hard to see this as a *conceptual* difficulty; see Bedau 1986.) This gives rise to a version of epiphenomenalism, according to which the mental is not causally relevant to the physical. The rejection of causal interactionism together with the obvious correlations between mental and physical events gave rise to parallelism, according to which mental and physical events evolve independently but in a way that gives rise to non-causal correlations, as the hands of two clocks, set independently a minute apart, may appear to be causally interacting because of the correlations in their positions, though they are not.<sup>37</sup> Parallelism is usually explained by reference to God's arranging things originally so that the mental and the physical develop in parallel (pre-established harmony), or through His constant intervention in bringing about what events, both physical and mental, give rise to the appearance of interaction (occasionalism).

Barring a reason to think that a property bearer cannot possess both irreducibly mental and physical properties, at most Descartes's arguments establish that there *could* be things which have only mental properties, as well as things which have only physical properties, not that there are or must be. If we can establish a priori at most that dualism could be true, whether it is true is to be determined, insofar as it can be, by empirical investigation. So far, there seems to be no very good empirical reason to suppose dualism is true.<sup>38</sup>

Idealism is the historical successor to dualism. It is dualism without material substance. Thus, it is a universal, pure mental particle theory. The classical position

is laid out in George Berkeley's *A Treatise Concerning the Principles of Human Knowledge* (1710). More sophisticated modern versions are called 'phenomenalism'.<sup>39</sup> Idealism is often motivated by a concern to understand the possibility of knowledge of objects of ordinary perception: forests and meadows, mountains and rain, stars and windowpanes. The Cartesian view of the relation of mind to world leaves it mysterious how we can have knowledge of it: if we know in the first instance only our conscious mental states, and whatever we can know by reason alone, yet the mental and material are conceptually independent, it looks as if we have no reason to believe that there is a material world causing our conscious experiences. Berkeley solved the problem by denying that objects of perception were material, and identifying them instead with collections of ideas (hence *idealism*). More recent treatments identify ordinary objects of common-sense knowledge with logical constructions out of phenomenal states. Berkeley denied also that we could even make sense of material substance. Leibniz (1714) likewise held that the basic constituents of things, monads (*unit*, from the Greek *monos*), were a sort of mind – though he did not hold that all were conscious – and that talk of ordinary things was to be understood in terms of monads and their states (as David Armstrong has put it, on Leibniz's view, "material objects are colonies of rudimentary souls" (1968, p. 5)). Kant (1781) is sometimes also interpreted as a phenomenalist. This view is not now widely embraced. It seems to be part of our conception of the world of which we think we have knowledge that it is independent of the existence of thinking beings, who are contingent players on the world stage.

Panpsychism holds that everything is a primary bearer of mental properties (not simply by being related to a primary bearer – as my chair has the property of being occupied by someone thinking about the mind-body problem). Panpsychism comes in reductive and non-reductive varieties. Its root can be traced back to antiquity (Annas 1992: 43–7). Panpsychists are represented among the Renaissance philosophers, and among prominent nineteenth-century philosophers, including Schopenhauer, W. K. Clifford, William James (at one time), and C. S. Peirce.<sup>40</sup> Panpsychism is associated often with (what seems to be) a revisionary metaphysics, with special motivations, as in the case of idealism, which is a reductive version of panpsychism. However, non-reductive panpsychism, which accepts a basic materialist ontology, is motivated by the thought that otherwise it would be inexplicable (a species of magic) that complex objects have mental properties. William James, in his monumental *Principles of Psychology* (1890), lays out this argument explicitly in chapter VI, "Evolutionary Psychology demands a Mind-dust." Thomas Nagel (1979a) has more recently revived the argument (see also Menzies 1988).<sup>41</sup> Panpsychism is a universal mental particle theory, and may be pure or mixed.

The double aspect theory should be thought of as a family of theories, rather than a single doctrine. What unifies the family is their affinity for being expressed with the slogan that the mental and the physical are different aspects by which we comprehend one and the same thing, though the slogan may be understood differently on different "versions" of the theory. Spinoza's doctrine of the parallelism

of thought and extension is the original of the double aspect theory, though he did not himself so describe his position.<sup>42</sup> Spinoza held that there was a single, infinite, eternal, and necessary substance, which had every possible categorical attribute, and so both extension and thought. Ordinary things were to be (re)conceived as modes (modifications) of the world substance. Thinking and extension were related in accordance with the parallelism thesis: “The order and connection of ideas is the same as the order and connection of things” (*Ethics*, IIP7; in 1994: 119–20). As Spinoza further explains it in the Scholium: “the thinking substance and the extended substance are one and the same substance, which is now comprehended under this attribute, now under that. So also a mode of extension and the idea of that mode are one and the same thing, but expressed in two ways” (ibid: 119). This is not an entirely pellucid doctrine. We understand it only to the extent that we understand Spinoza’s metaphysics, itself a matter of interpretive difficulty. The idea that the mental and the physical are two ways of comprehending one thing, however, can survive the rejection of Spinoza’s metaphysics, and has inspired a number of views which appeal to similar language.

If we allow a multitude of substances, the double aspect theory holds that every object, or some, can be viewed as mental or physical, depending on how we take it. In G. H. Lewes’s image (1877; repr. in Vesey 1964: 155), to comprehend a thing as mental or physical is like seeing a line as concave or convex: “The curve has at every point this contrast of convex and concave, and yet is the identical line throughout.” The double aspect theory is not currently popular. Partly this is due to its unclarity. It is intended to be more than the claim that there are objects that have mental and physical properties, neither being conceptually reducible to the other (though sometimes it has been used in this broader sense), or even that there are systematic correlations between everything physical and something mental.<sup>43</sup> But there seems to be nothing more in general to say about what it comes to, and we must rather look to particular theories to give it content. Its lack of popularity is partly due to factors independent of the details, and, in particular, to the dominance of our current scientific worldview, according to which the world once contained no thinking things, and has evolved to its present state by natural law.

Double aspect theories may be either universal or restricted, mixed mental particle theories. Some double aspect theories are versions of panpsychism, then, as in the case of Spinoza, since he does maintain that everything has mental properties. Compatibly with the guiding idea, however, one might also maintain that some objects have two aspects, two ways of comprehending them, mental and physical, though not all do.<sup>44</sup>

Finally, there is what I call the special particle theory, which holds that some basic constituents of things, which are at least spatially located, have mental properties, but not all. This counts as a restricted, mixed mental particle theory, counting spatial location as a broadly physical property. So far as I know, this is not a view that has been represented among traditional responses to the mind–body problem.<sup>45</sup>

### 1.6.2 *Conceptual anti-reductionism*

Rejecting proposition (3) leads to emergentism. There are in principle two varieties, neutral emergentism and emergent materialism, according to whether basic constituents are conceived as physical or neither physical nor mental. Most emergentists are materialists, and I concentrate therefore on emergent materialism. Emergent materialists hold that there are only material things, but that some complex material things, though no simple ones considered independently of complexes in which they participate, have mental properties, and that those mental properties are not conceptually reducible to any of the physical properties of the complexes that have them. Emergentism historically was a response to the rejection of forms of dualism and idealism in favor of a materialist ontology. It is associated with the rise of science generally in the nineteenth century, and the development of the theory of evolution in particular. It dispenses with the ontological, but retains the conceptual anti-reductionism of Cartesian dualism. Late nineteenth- and early twentieth-century emergentists included T. H. Huxley (“Darwin’s bulldog”; 1901), Samuel Alexander (1920), C. Lloyd Morgan (1923), and C. D. Broad (1925). The term “emergent” was pressed into service because the universe was thought to have once not contained any objects that had any mental properties. Since all its objects are material objects, once they had no mental properties, but now some do, and those properties are not conceptually reducible to physical properties, mental properties must emerge from, in some way, certain organizations of matter, though this cannot be deduced from a complete description of the objects that have mental properties in terms of their physical properties.<sup>46</sup> Emergentists take seriously the evidence that at least some aspects of the mental are not in any sense physical phenomena. This was the traditional view, and is undeniably an initially attractive position. Once we have extricated ourselves from the confusions that lead to the view that there must be mental substances distinct from material substances to bear irreducible mental properties, the view that we are latecomers to the physical world – natural objects that arose by natural processes from materials themselves falling wholly within the realm of mechanics – leads naturally to emergent materialism.

Varieties of emergentism arise from different views about the relation between fundamental properties and mental properties. Traditional emergent materialists held that there were type-type nomic correlations between physical and mental states. This is to hold that for every mental property some sentence of the form [C] obtains with the force of nomological necessity. One may hold that mental properties merely nomically supervene on physical properties, and that there are no type-type correlations.<sup>47</sup> Finally, one might hold a version of what is called ‘anomalous monism’. Anomalous monism was originally proposed as a thesis about the relation of mental and physical events (Davidson 1980). It holds that every mental event is token identical<sup>48</sup> with a physical event, but there are no

strict psychophysical laws, and so no strict bridge laws.<sup>49</sup> This still allows loose, *non-strict*, nomic supervenience or nomic type correlation. A stronger version denies even that there are loose nomic relations between mental and physical event types. The idea can be adapted to objects as the view that though some complex objects have mental properties, there are no strict nomic correlations or supervenience relations between physical and mental properties, or, in the stronger version, none at all.

Emergentism is often (nowadays especially) associated with epiphenomenalism.<sup>50</sup> Epiphenomenalism holds that mental properties are not causally relevant to anything (or, at least, to anything physical). Among late nineteenth- and early twentieth-century emergentists there was disagreement about the causal efficacy of the mental. Some (e.g. Morgan and Broad) held that there were not only emergent properties, but also emergent laws governing systems at the level of the emergent properties which could then affect the course of events at lower levels (downward causation).<sup>51</sup> This stream in the emergentist tradition has now nearly run dry (though see Sperry 1986).<sup>52</sup> Other prominent emergentists saw the mental as wholly dependent on the physical, and causally inert. In a famous discussion, T. H. Huxley held that consciousness was “the direct function of material changes” (1874: 141), but also that consciousness was as completely without power to affect the movements of our bodies “as the steam-whistle which accompanies the working of a locomotive engine is without influence upon its machinery” (p. 140). (See also Hodgson 1870; G. J. Romanes 1895.) On this view, mental activity is a shadow cast by neural activity, determined by it, but determining nothing in turn: conscious mental states are “nomological danglers,” in Feigl’s apt phrase (1958).

Until the second half of the twentieth century, emergentists believed that there were type-type correlations between the states of our central nervous systems and mental states that held as a matter of natural law. These laws were not purely physical, but bridge laws, since their statement involved irreducibly both mental and physical predicates. Epiphenomenalism is motivated by the thought that the universe would proceed just as it has physically if we were simply to subtract from it the bridge laws: we do not need in principle to refer to any non-physical events or laws to explain any physical event. Just as the locomotive would continue in its path if we were to remove its whistle, so our bodies would continue in their trajectories if we were to remove their souls.<sup>53</sup> The conjunction of the view that there are such type-type nomic correlations, and the view that the physical is a closed system, is nomological reductionism. Obviously, the further we move from nomic type-type correlations, the less plausible it becomes that we can find a place for the causal efficacy of mental properties. The perceived threat of epiphenomenalism has been one of the motivations for physicalism. It is an irony that some popular ways of trying to ground physicalism also raise difficulties for seeing how mental properties could be causally relevant to what they are supposed to be.<sup>54</sup>

### 1.6.3 *Conceptual reduction*

To reject proposition (2) is to adopt conceptual reductionism for mental properties.

We consider first, briefly, non-physicalist ways of rejecting (2). There are two possibilities: that the mental is conceptually reducible to, or supervenes on something non-physical. While the latter position is an option, it has not been occupied. However, neutral monism, the view that the mental and the physical might both be understood in terms of something more basic, enjoyed a brief run at the end of the nineteenth and in the first half of the twentieth century.<sup>55</sup> The view is associated with William James (1904), who argued that “pure experience” is the primal stuff of the world and minds and objects were to be conceived of as different sets of experiences, so that the same experience could be taken with one set as a thought, and with another as a component of an object thought about. Neutral monism, as advocated by James, rejects the view that there is a subject of experience, and retains only what was traditionally thought of as its object. As James put it, “those who cling to it are clinging to a mere echo, the faint rumor left behind by the disappearing ‘soul’ upon the air of philosophy” (pp. 3–4). Ernest Mach (1886) held a similar view, and Bertrand Russell developed a version of neutral monism, inspired by James, in which sensibilia (or “sensations” as Russell put it in *The Analysis of Mind* (1921)), introduced originally as mind-independent objects of direct awareness (1917), played the role of the neutral stuff out of which minds and physical objects were to be logically constructed (1921).

It may seem as if this view should more properly be described as a version of idealism, because the terms that James, Mach, and Russell used to describe the neutral stuff are usually associated with mental phenomena. But they held that the neutral stuff was not properly thought of as mental in character, but only when it was considered in a certain arrangement. It might then seem reasonable to describe neutral monism as a double aspect theory, at least in the sense that it treats each of the fundamental things as a thing that could participate in a series of things which constituted something mental, as well as in a series of things which constituted something physical; thus, each could be said to be viewed under a physical or a mental aspect. However, since talk of thoughts and material things is conceived of as translatable into talk neither mental nor physical, neither the mental nor the physical has a fundamental status in the ontology of neutral monism.<sup>56</sup> Rather, both bear the relation to the neutral stuff that ordinary objects do to phenomenal experience according to idealist theories. Just as idealist theories do not countenance genuine material substance, neutral monism does not countenance genuine mental or physical substances in its fundamental ontology, though it gives an account of *talk* of each sort.

Neutral monism has some theoretical virtues. It avoids the difficulties associated with trying to reduce either the mental to the physical or vice versa, and, if successful, provides a fundamental, unified account of things of all kinds in terms

of a fundamental kind, the dream of idealists and physicalists alike. Despite this, it is not a popular view. It attracts neither those who think the mental is a basic feature of reality, nor those who dream of the desert landscape of physics. Moreover, it is difficult to develop the account in detail, and difficult to understand the nature of the neutral stuff which it relies upon.

We turn now to physicalist rejections of proposition (2).

The first twentieth-century physicalist position to gain popularity was logical behaviorism, which was spurred on in part by the verificationism of the logical positivists before the Second World War, the view that the meaning of a sentence was to be sought in the empirical conditions for confirming or disconfirming it (a view with roots in classical British empiricism).<sup>57</sup> Logical behaviorism has a stronger and a weaker form. The strong form I will call ‘translational behaviorism’, and the weaker form ‘criterial behaviorism’. Translational behaviorism holds that every psychological statement can be translated into a statement about actual and potential behavior of bodies. Criterial behaviorism holds, in contrast, merely that there are behavioral analytically sufficient conditions for the application of mental predicates.

Logical behaviorism has long fallen out of fashion. This is explained in part by the fall from favor of verificationism, which provided it theoretical support, but also by the fact that not only were no satisfactory translation schemes advanced, but there are reasons to think none could be forthcoming in principle. A particularly troubling problem was that what behavioral manifestations we may expect from someone with a certain mental state depends on what other mental states he has. Consequently, there can be no piecemeal translation of psychological claims into behavioral terms. In addition, behaviorism seems incompatible with our conception of mental states as (possible) causes of behavior. For to reduce talk of mental states to talk of behavior is to treat it as merely a more compendious way of describing behavior. Behavior, though, cannot cause itself.<sup>58</sup>

The two principal physicalist responses to the defects of behaviorism were analytic functionalism and the psychophysical identity theory. Though the psychophysical identity theory came to prominence before analytic functionalism, it will be useful to discuss functionalism first, since it is the natural successor to logical behaviorism, and this will put us in a position to usefully clarify the psychophysical identity theory, which in some early versions suffered from a number of confusions and conflicting tendencies.

Analytic functionalism holds that mental states are conceptually reducible to functional states. Functional states are held to conceptually supervene, in the sense defined in section 1.4, on physical states.<sup>59</sup> The identification of mental with functional states then leads to *physicalism* without conceptual reduction of the mental to the physical per se. A functional state, in the relevant sense, is a state of an object defined in terms of its relations to input to a system, other functional states of the system, and output from the system. Some of the logical behaviorists, e.g., Gilbert Ryle in *The Concept of Mind* (1949), can be seen to have been moving toward something like this (functionalism may therefore be said to be the

eclosion of behaviorism). Functionalism was inspired, at least in part, by the rise of computer technology<sup>60</sup> after the Second World War. Its earliest form in the twentieth century, machine table functionalism, introduced by Hilary Putnam (1967), was directly inspired by theoretical work on finite state machines, which is what a (finite state) computer is.<sup>61</sup> A machine table describes a system in terms of a list of exhaustive and mutually exclusive inputs, a list of possible states, a list of outputs, and, for each possible state, what state it moves to and what output is produced given that it receives a given input. The operation of any computer running a program can be described exhaustively in terms of a machine table. For programmable computers, the program determines what machine table it instantiates (relative to a division of a system into states of particular interest to us). Putnam generalized the notion of a finite state automaton (a system describable using a finite state machine table with deterministic state transitions) to a probabilistic finite state automaton, in which transitions are probabilistic. The general form of the proposal is that a system is in a certain mental state iff it has an appropriate machine table description and appropriate inputs or appropriate states. Putnam treated his proposal as an *empirical* hypothesis. This is typically called ‘psychofunctionalism’, following Block (1978).<sup>62</sup> It is nonetheless one of the principal inspirations for analytic functionalism, and is easily reconstrued as a thesis about our concepts of mental states. *Theoretical* or, sometimes, *causal role* functionalism is a variant on the theme. On this view, we start with a theory that embeds psychological terms. The concepts expressed by these terms are taken to be concepts of states that are characterized exhaustively by their relations to other states and inputs and outputs as specified abstractly in the theory.<sup>63</sup>

Functionalism is attractive. It accommodates a thought that motivated behaviorism, namely, that our mental states are intimately tied up with understanding of behavior, but it does so in a way that distinguishes them from, and treats them as causes of, behavior. Moreover, functionalism allows for the possibility of immaterial thinking beings, since a system’s having a certain functional organization does not depend on what it is made of, but rather on its causal powers with respect to inputs and outputs. It has merely to sustain the right organization mediating inputs and outputs. Functional states are *multiply realizable*. This accommodates one of the thought experiments that motivates the assumption of the conceptual independence of the mental and the physical. It finds a place for the mental in the natural world that exhibits it as grounded in the physical, in the sense that it exhibits the mental as conceptually supervening on the physical, without insisting on a conceptual reduction to physical properties. It thereby allows that the language of psychology is distinct from that of physics, while allowing that the realization of psychological states requires nothing more than objects having physical properties governed by physical laws. The multiple realizability of functional states also (prima facie) protects functionalism from a charge leveled against the psychophysical identity theory, namely, that it would be implausible, and chauvinistic, to insist that only those physically like us can have mental states.<sup>64</sup>

Analytic functionalism has come in for considerable criticism, but remains popular, especially outside philosophy in fields contributing to the new discipline of cognitive science. A first objection to functionalism is that no one has come up with a successful conceptual reduction of mental concepts to functional concepts. It might be said that this could equally well be a sign of the complexity of these functional concepts. A second objection to functionalism is based on the *prima facie* intelligibility of systems which are functionally identical to us but which have no mental states. An example is provided by a thought experiment of Ned Block's (1978).<sup>65</sup> Imagine a robot body actuated by a program instantiating a machine table for some person. Imagine further that we instantiate the program by providing each member of the population of China with a two-way radio with a display that shows the current input to the robotic system and an indicator of whether the system is in his state. Each person presses a button on the radio appropriate for the input when his state is active. Signals are relayed to the body for appropriate action. Suppose that the Chinese get so good at this that our robot and accessories constitute a system functionally identical to our original. Does this system now constitute an intelligent, conscious being? Most people, first confronted with the thought experiment, deny that we have created a new person (who will die when the exercise is terminated).<sup>66</sup>

Another important objection is also due to Ned Block (1978). Functionalists must decide how to specify inputs and outputs to the system. This presents them with a dilemma. If we specify the inputs and outputs physically using ourselves as models, it is not difficult to describe some system that could have a mind that is incapable of causing those outputs, but causes others instead (e.g., we do not want to rule out, *a priori*, intelligent jellyfish, or beings whose inputs and outputs are various portions of the electromagnetic spectrum, and so on). Further, it is difficult to see how we could put *a priori* limits on the physical character of inputs and outputs. However, if the inputs and outputs are specified barely as distinct, then it is not unlikely that we can find minds just about everywhere, for it is plausible that most complex systems will admit of some division into states and inputs and outputs that will instantiate some machine table said to be sufficient for having a mind (e.g., the world economy).

It also has been objected that it is easy to imagine functional duplicates who differ in the qualities of their experiences. A well-known thought experiment designed to show this is that of the inverted spectrum. We imagine two individuals functionally indistinguishable, and therefore behaviorally indistinguishable, but imagine that their experiences of the colors of objects in their environments are inverted with respect to one another. Where one experiences a red object, e.g., the other experiences a green object. They both utter the same sentence in describing it, but each sees it differently. If this is conceivable, then their color experiences are not conceptually reducible to their functional organization, and, hence, functionalism is false with respect to these phenomenal qualities.<sup>67</sup>

Another difficulty is that it is unclear that functional states can be causally relevant to the right sorts of behavior. Functionalism accommodates mental states

as *causes* of behavior by definition.<sup>68</sup> But this may secure the causal connection in the wrong way. For a state defined in terms of its effects in various circumstances cannot be the type in virtue of which those effects come about. Causal relations between events or states are underlain by *contingent* causal laws connecting types under which they fall.<sup>69</sup> One type is causally relevant to another type (in certain circumstances) iff they are connected by a causal law (in the circumstances). However, the relation between a functional state and the output (type) in terms of which it is partially defined is not contingent. Thus, the state type and output type cannot feature appropriately in a contingent causal law. Therefore, functional state types are not causally relevant to output in terms of which they are defined.<sup>70</sup> If this reasoning is correct, analytic functionalism entails epiphenomenalism with respect to these outputs. An advantage of functionalism over behaviorism was supposed to be that it makes mental states causes of behavior. The trouble is that it does so in a way that undercuts the possibility of those states being causally relevant to what we expect them to be.

Worse, it seems quite plausible that we do conceive of our mental states as causally relevant to the behavior that we would use to define mental states on a functional analysis. Our beliefs about the causal relevance of mental states to behavior may be false. It is contingent on what causal laws hold. But if they are not necessarily false, then functionalism cannot be true, since it precludes *the possibility* of our mental states being causally relevant to our behavior.<sup>71</sup>

Let us now turn to the psychophysical identity theory. This is the view that mental properties are physical properties. I start with what I believe is the most plausible form of the psychophysical identity theory, which is based on an approach advocated by David Lewis (1966, 1972). The approach makes use of functionalist *descriptions* of states extracted from a “folk theory” of psychology to identify mental states with physical states.

Analytic functionalism holds that psychological concepts and properties *are* functional concepts and properties. This should be distinguished from the view that psychological properties are *picked out by functional descriptions*. This view does not reduce mental properties to functional properties. Rather, it treats mental terms as theoretical terms. Theoretical terms are treated as picking out properties in the world (and so as expressing whatever concepts are of those properties) that actually play the role the theory accords them in the systems to which it is applied. We represent our psychological theory as a single sentence, ‘ $T(M_1, M_2, \dots, M_n)$ ’, where ‘ $M_1$ ’ and so on represent psychological terms referring to properties. Then we replace each such term with a corresponding variable, ‘ $x_1$ ’, ‘ $x_2$ ’, and so on, and preface the whole with a quantifier for each, ‘there is a unique  $x_1$  such that’ (symbolized as ‘ $(\exists!x_1)$ ’), etc., to yield, ‘ $(\exists!x_1)(\exists!x_2) \dots (\exists!x_n)T(x_1, x_2, \dots, x_n)$ ’. The property “ $M_1$ ” picks out can be characterized as follows, where we leave out the quantifier in front of ‘ $T(\dots)$ ’ associated with ‘ $x_1$ ’:

$M_1$  is the unique property  $x_1$  such that  $(\exists!x_2) \dots (\exists!x_n)T(x_1, x_2, \dots, x_n)$

In application to human beings, on the assumption that the theoretical description of this property is satisfied by a physical property of our bodies or central nervous systems, it follows that  $M_1$  *is* that physical property. Thus, we arrive at a psychophysical identity theory.

Given how we have characterized the relation between concepts, predicates, states, and properties, if we identify a mental state or property with a physical state or property, it follows that the corresponding mental concept *is* a physical concept. Therefore, the view that mental properties are picked out by functional descriptions will lead to the conclusion that mental concepts are conceptually reducible to physical concepts, if those descriptions pick out physical states or properties.<sup>72</sup> This is not, however, something we could know a priori. It could only emerge after empirical investigation. For on this view, the concepts expressed by our theoretical terms are hostage to the nature of the phenomena to which we apply them. We start only with descriptions of the properties, and so, in effect, only with descriptions of the concepts of them. We can reason a priori using the concepts only after we have discovered them a posteriori.

The psychophysical identity theory has the advantage over functionalism and emergentism in securing the causal relevance of mental properties. No one doubts that our physical states are causally relevant to our movements. Identifying mental states with physical states, the psychophysical identity theory makes their causal relevance unproblematic. Some philosophers have argued that since only identifying mental with physical states will secure their causal efficacy, and mental states are causally efficacious, we are justified in identifying them (Papineau 1998).

This comes at a cost, though. On this view, prior to empirical investigation it is open that there are no mental properties at all, no properties that answer to the theoretical descriptions we have of them. This shows that this view has in common with eliminativism the assumption that we do not know directly that anything has the properties we suppose to be picked out by our psychological terms. A view like this entails eliminativism when combined with the claim that no physical (or any other) states play the required roles. To the extent to which we find it implausible, perhaps even unintelligible, that we could discover we don't have any mental states, we should find equally implausible or unintelligible the argument for the psychophysical identity theory just reviewed.<sup>73</sup>

The psychophysical identity theory (also called “central state materialism”), like functionalism, has antecedents that stretch back to the ancient world. In the twentieth century, it was influentially advocated after the Second World War by Ullin Place (1956), Herbert Feigl (1958), and J. J. C. Smart (1959).<sup>74</sup> Place and Smart held that sensations were to be theoretically identified with brain processes, in the same way that lightning was identified with a certain sort of electrical discharge (this can be generalized straightforwardly to states; see Armstrong 1968).<sup>75</sup> They thought of this as a contingent identity, because it was empirically discovered. The position is also sometimes called ‘the topic neutral approach’, because Smart in particular argued that in order that we not have irreducible mental properties, and yet make sense of the possibility of contingent identity, the descriptions by which

we pick out mental processes (more generally mental states), which are to be empirically identified with physical ones, must leave it open whether they are physical or not. This position came into considerable criticism for the claim that identities could be contingent (see Kripke 1980: 98–100, 144–55). If we are speaking about strict identity of things – in the present case, properties – there is no room for contingency, since identity holds of necessity between everything and itself, and between no distinct things. The view I have presented based on Lewis’s approach is a descendant of these early psychophysical identity theories. It retains the view that mental properties are physical properties (on the assumption that unique physical properties play the right roles). But it rejects the view that this is contingent (given that in fact there are physical properties playing the right roles). Seeing theoretical terms as introduced to track properties that are to play certain roles helps us to see how the discovery of identities can be empirical although the identities are necessary. It also gives precise content to the idea that the descriptions that pick out mental states are topic neutral, since they are to be given by the structure induced by our folk theory of psychology.

At this point, a note on metaphysical necessity is in order. This modality is often invoked in contemporary discussions of the mind–body problem. It is said to be distinct both from nomological and conceptual necessity, stronger than the former, and weaker than the latter. How did it come to be introduced? A paradigm of metaphysical necessity is supposed to be the sort that results from theoretical identifications involving *natural kinds*, like the identification of gold with that element with atomic number 79. It is not contingent or just a matter of natural law, but necessary that gold is the element with atomic number 79, since nothing that did not have atomic number 79 would count as gold even in a world with different natural laws. Still, it was an empirical discovery, and not something we could have known purely a priori. But since conceptual truths are knowable a priori, it must be that metaphysical necessity is distinct from conceptual necessity – or so the argument goes.

The perceived utility of metaphysical necessity is that it provides a way to argue for connections between the mental and the physical stronger than nomological connections, indeed, identities, which at the same time is immune to refutation by thought experiments that seem to show mental and physical phenomena are independent. Since metaphysical necessity is supposed not to be governed by what is conceptually possible, and such thought experiments are, they fail to bear on the claim.<sup>76</sup>

As I said earlier, in my view no philosopher has succeeded in expressing a concept by ‘metaphysical necessity’ that answers to this argument. The first thing that should make us suspicious about “metaphysical necessity” is that we do not have any account of what grounds claims supposedly about it. Barring this, it is dubious that we have any precise idea of what is supposed to be expressed here by the term ‘metaphysical’. The second thing that should make us suspicious is that there is available a straightforward explanation of the facts which motivate introducing metaphysical necessity that requires no mysterious new sort of necessity.

Our reading of Lewis’s account of theoretical identifications provides the key. On that account, we associate with each theoretical term a description of the property that it picks out (the property *P* which plays such and such a role in such and such systems). It is a matter for empirical investigation what property actually satisfies the description (as it is in determining which individual is the mayor of New York). However, the concept a term expresses is, as we have seen, what determines the property it picks out: they are a matched set. Thus, to discover what property a theoretical term picks out by discovering empirically what satisfies the associated description is likewise to discover empirically what concept the term expresses. Prior to that, we had a description of a concept, but it was not given to us directly. Thus, when we discover that ‘is gold’ picks out the element with atomic number 79, we discover what concept it expresses. Prior to this, we did not know what concept it expressed. Once we know, we are in a position to see that ‘Gold is that element with atomic number 79’ expresses a conceptual truth, which is knowable a priori. What was not knowable a priori was not that gold is that element with atomic number 79, but that ‘gold’ expressed the concept of the element with atomic number 79. We competently use such natural kind terms prior to discovering what concepts they express. This is explained by the fact that we treat such terms as tracking properties that *explain easily identifiable features of things* we in practice apply them to. We apply the terms in accordance with those features. The mistake in the original argument was to confuse competence in applying natural kind terms with grasp of the concept expressed: given that we do not know what property is picked out, we likewise do not know what concept is expressed. What we know is just what work the property is supposed to do, which enables us to develop an application practice with the term that is to pick it out.

Thus, the introduction of ‘metaphysical necessity’ is gratuitous. We have no reason to suppose anything corresponds to it, and no idea of what it would be if it did. Consequently, we cannot look to metaphysical necessity for new avenues for the solution of the mind–body problem.<sup>77</sup>

Before we leave the topic of reductionism, it is important to consider a hybrid view that combines functionalism and externalism about thought content. Externalist accounts of mental states emphasize the importance of our relations to things in our environments in conceptually individuating them. At the same time that difficulties were mounting for functionalism, independently some influential arguments were advanced which suggested that content properties were relational properties.<sup>78</sup> According to these accounts, what thoughts we have depends on what actual and potential causal relations we bear to things in our environments. (Relationally individuated states are often called ‘wide states’ in the literature, and non-relationally individuated states ‘narrow states’.) The most important division among externalist views is that between physical and social externalism. Physical externalism holds that thought contents are individuated (in part) by relations to our physical environments. Social externalism holds that thought contents are individuated (in part) by how others in our linguistic communities

use the words we intend to use as they do.<sup>79</sup> A reductionist externalist account of thought content will typically hold that our concepts at least of contentful mental states can be reduced to functional and causal concepts, where we include systematic causal relations to external things in fixing the contents of thoughts.

Externalist theories too have come in for considerable criticism. Two are worth mentioning because they are connected with themes already touched on.<sup>80</sup> The first is the objection that if externalism were true, we would not be able to know the contents of our own thoughts without empirical investigation, but since we must in order to undertake empirical investigations in the first place, externalism entails unacceptably that we can never know the contents of our own thoughts.<sup>81</sup> The second is connected with a difficulty already noted for functionalism. It is that treating content properties as individuated in part in terms of relational properties threatens to make them unsuitable for explaining our behavior (described physically). The problem is not that relational properties cannot be causally relevant to anything. There are *prima facie* counterexamples to this. That something is a planet, for example, may be cited in explaining why I come to believe that it is. But the difficulty for externalism only requires that the kind of relational properties that content properties would turn out to be could not be causally relevant to our behavior. For externalist theories exploit the possibility of behavior (described physically) remaining the same because one's non-relational physical states remain the same while one's thought contents vary. This appears to show that the relational states are "screened off" from the relevant effect types by the non-relational physical states, which are sufficient to account for the behavior and are independently necessary.<sup>82</sup>

The conception of our (at least conscious) mental states as of a sort which are (a) non-inferentially knowable by their possessor (our concepts of which are therefore not theoretical concepts), though by no one else, and (b) as (possibly) causally relevant to other sorts of things (other mental events and states as well as non-mental events and states) may be called the core of the Cartesian conception of the mind. The difficulties we have been reviewing for reductionist proposals about the mental are connected with these features. No physical states seem capable of possessing both. The first feature stands in the way of the plausibility of the psychophysical identity theory, and, arguably, of externalism about thought content. The second seems to preclude conceptual reduction to states characterized in terms of their causal relations to other things, or, again, in terms of their relations to things in the environment.

#### 1.6.4 *Irrealism*

Finally, we turn to eliminativism. Eliminativists seek absolution through denial. According to eliminativism, nothing has mental properties. Prominent proponents of this position are Paul Churchland (1981) and Stephen Stich (1983), who argue that our mental concepts are empty.<sup>83</sup> They are concepts deployed in

a pre-scientific or “folk” theory of behavior, which are ripe for replacement by a more sophisticated theory deploying different categories, which answer better to our explanatory interests. Folk psychology goes the way of theories of disease that appeal to demonic spirits. The psychological entities of our common-sense conceptual scheme too are creatures of darkness. We must now march forward into a brighter future, out from under the shadow cast by superstitions inculcated in the childhood of civilization, shriven of the sin of belief in the mind.

Eliminativism remains, not surprisingly, a minority position. It has some advantages – as Karl Popper has said, “the difficult body–mind problem simply disappears, which is no doubt very convenient: it saves us the trouble of solving it” (1994: 8). But it is hard to credit. It must reject the view that knowledge of our own conscious mental states is epistemically prior to knowledge of other things, which seems to be in conflict with a very natural account of how we come to know things about the world around us through perceptual experience. There are also certain difficulties involved in thinking about our position in putting forward the theory, and in accounting for how we could justify it. For surely if someone maintains that the theory is true, there is at least one person who believes something, namely, that eliminativism is true, in which case, eliminativism is false. The difficulty is that we have no vocabulary for describing the acceptance, rejection, and support of theories that does not presuppose that theoreticians have mental states. Eliminativists maintain this is merely a pragmatic difficulty, but it is not one that they have overcome.

## 1.7 Conclusion

This concludes our survey of the mind–body problem and the principal responses to it. A summary of the positions we have considered is given in figure 1.1.

Two basic positions mark the continental divide of the mind–body problem. All the positions we have examined are expressions of one or the other of them. One accepts the mental as a basic feature of reality, not explicable in terms of other features. Its basic characteristic is that it accepts propositions (1) and (2), realism and conceptual autonomy. The other insists that the appearance that the mental is a basic feature of reality must be an illusion, and that we and all our properties can be understood exhaustively ultimately in terms that make intelligible to us at the same time the clearly non-mental phenomena of the world. Its basic characteristic is that it accepts propositions (3) and (4), constituent explanatory sufficiency and constituent non-mentalism. The second view, constrained by the assumption that the basic constituents of things are physical (constituent physicalism), is equivalent to physicalism, with eliminativism as a degenerate case. The reason the mind–body problem does not go away, despite our being clear about the options in responding to it, is because of the constant battle between common sense, which favors the view that the mental is a basic feature of reality,

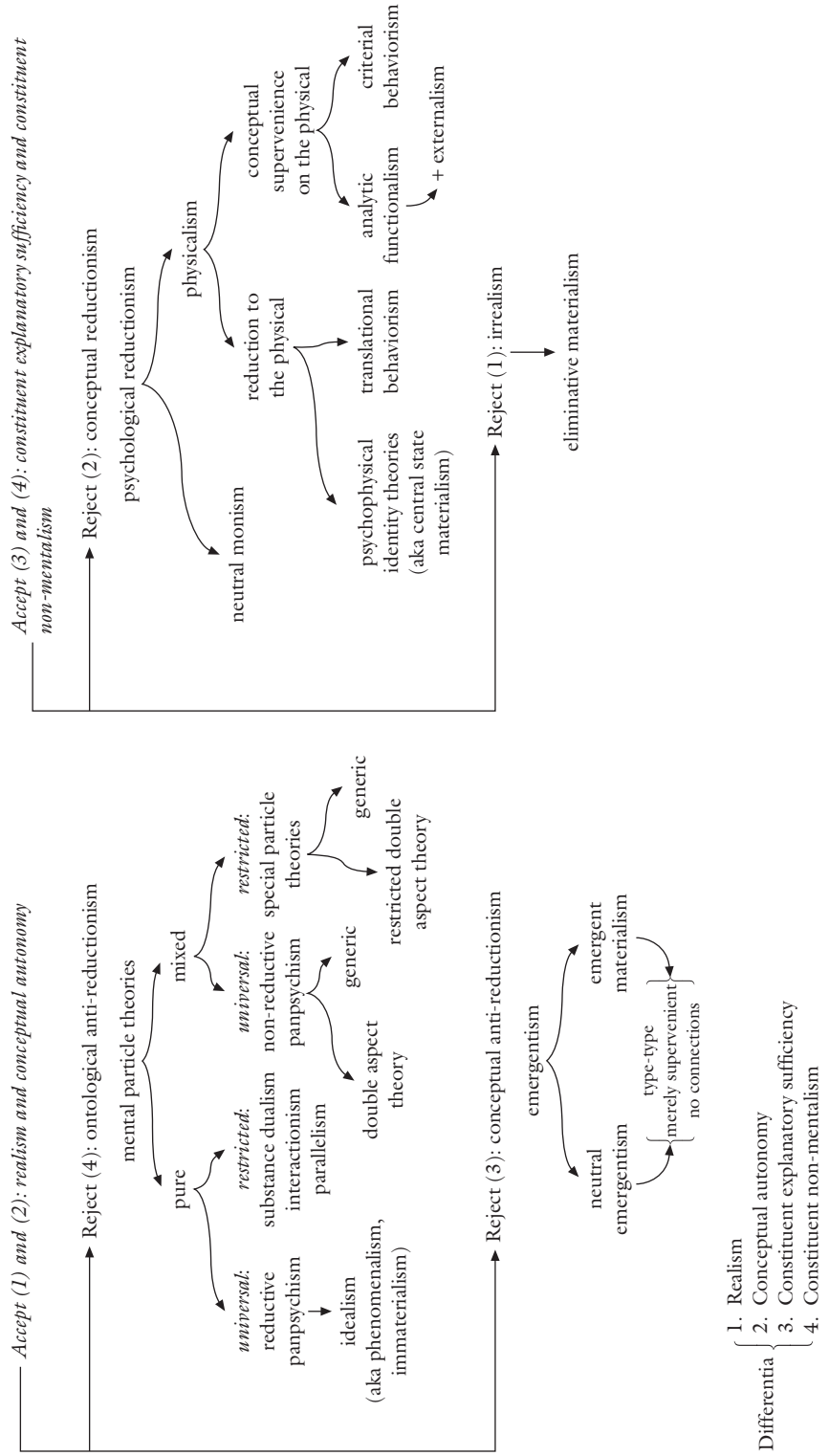


Figure 1.1 The logical space of solutions to the mind-body problem

and the pull to see it as an authoritative deliverance of science that this is not so. We find ourselves constantly pulled between these two poles, unable to see our minds as nothing over and above the physical, unwilling to see the universe as containing anything not explicable ultimately in terms of its basic, apparently non-mental, constituents.

### Notes

- 1 The term ‘the mind–body problem’ is not used univocally. What guides my usage is an interest in getting at the puzzle that has generated the great variety of positions that we find in philosophical and scientific discussions of the relation of mental phenomena to physical phenomena. If I am right, there is a puzzle we can articulate clearly to which all the positions on the relation of the mental to the physical can be seen as responses. If any one problem deserves the label ‘the mind–body problem’ it is this.
- 2 In the course of discussion, a considerable amount of terminology will be introduced. This is partly to enable us to state our problem and its possible solutions with precision. More terminology is introduced than is strictly necessary for this. The excess is intended to provide a foundation for further reading in the relevant literature on the topic. I will often provide references representative of particular views or arguments. I list here some collections of papers which together give a fairly comprehensive picture of the historical and contemporary development of views on the mind–body problem: Vesey (1964), Anderson (1964), O’Connor (1969), Borst (1970), Rosenthal (1971), Block (1980), Eccles (1985), Lycan (1990), Rosenthal (1991), Beakley and Ludlow (1992), Warner and Szubka (1994), Block et al. (1997), Cooney (2000). Rosenthal (1991) is particularly comprehensive. Vesey (1964) contains historical sources not found in the others. Anderson (1964) contains early papers on the computer model of the mind. Eccles (1985) contains contributions mostly by scientists, both philosophical and scientific in character. Block et al. (1997) is devoted specifically to recent work on consciousness.
- 3 “By the term ‘thought’,” Descartes says, “I understand everything which we are aware of as happening within us, in so far as we have awareness of it” (1984, vol. I: 195 [1644: I.9]). This corresponds to the feature of consciousness I describe below as non-inferential knowledge of our modes of consciousness. Descartes held also that a state is a mental state *only if* it is conscious, but this is widely regarded as too stringent a requirement, for reasons considered below.
- 4 On this common-sense conception of events as changes, they are datable particulars. They may be complex as well as simple. My snapping my fingers is an event. So was the Second World War. If an object changes from being F to being non-F, the event is the changing from being F to being non-F. If we individuate events in terms of which objects, times, and properties they are changes with respect to, the question whether mental events are physical events is reduced to the question whether mental properties are physical properties.
- 5 It is sometimes thought that this is too strong. For one might mistakenly think, e.g., that one is in pain because one expects to be, given the occurrence of some event one had anticipated and expected to cause pain. For example, someone might think he was in pain when someone puts an ice cube on the back of his neck, if he had been

told that a piece of metal heated red hot was about to be pressed against the back of his neck. The possibility of his having a false belief in these circumstances does not show, however, that he does not know what he experienced. For he will correct his mistake. He will realize quickly that he is not, and was not, in pain. For he can recall what the experience was like. That requires knowing what character it had at the time, since one cannot remember something one did not originally know. Memory preserves but does not create knowledge.

- 6 For discussion of this issue, see essays 20–24 in Block et al. (1997).
- 7 See Nagel (1979b, 1994, 1998), and McGinn (1989, 1991, 1999). McGinn and Nagel think there must be a way of understanding how the operations of our brains give rise to consciousness, but that we currently have no conception of how that could be. McGinn is the more pessimistic, since he thinks whatever the correct explanation, it is one that we cannot in principle understand, given our cognitive make-up, while Nagel thinks we may one day develop appropriate concepts. The view that consciousness is the central difficulty is as old as discussion of the mind–body problem.
- 8 This terminology traces back to medieval philosophy; it is derived from the Latin verb *intendere*, for ‘point at’ or ‘aim at’; it was used to characterize the object of a thought when it did not exist *in reality*, but had *intentional inexistence*, or *existed only intentionally in the thinking subject*.
- 9 Some things besides attitudes of the sorts we have been discussing can be said to represent things, and so to have intentionality; e.g., a sentence, or a portrait. However, these have representational content only because agents treat them as representations in accordance with various rules. This is derived, as opposed to original, intentionality (Searle 1983, 1984). Mental states have original intentionality. I use ‘intentionality’, without qualification, to mean original intentionality.
- 10 A *disposition* is a state of an object that consists in its settled tendency to undergo some change in certain conditions. Water solubility is a simple dispositional state possessed by salt and sugar: when placed in unsaturated water in a certain range of temperatures and pressures, they dissolve. The change undergone that characterizes a disposition is its *manifestation property*, the property that is manifested. The *manifestation condition* is that under which the manifestation property is manifested. Often both of these are encoded in the name of the disposition, as in “water solubility.” Dispositional attitudes are not simple dispositions, but what Gilbert Ryle called “multi-track dispositions” (1949: 43–4). This means that they manifest themselves in various conditions in various ways. Moreover, they are interlocking dispositions: among the manifestation conditions for any given attitude will be conditions involving what other attitudes an agent has. A desire to buy a certain book will not be manifested unless I *believe* I have the opportunity to purchase it, and have no other desires whose satisfaction I rank above that for the purchase of the book, and which I think I can satisfy only to its exclusion.
- 11 Many recent theories of cognitive activity have appealed to in principle unconscious inferences in their explanations, thereby presupposing the two can be conceived independently. See Ludwig (1996c) for criticism of these views.
- 12 Some philosophers have recently argued that conscious states may be exhaustively characterized in terms of their representational content. Examples are Lycan (1996), Dretske (1997), and Tye (1997). For contrary views, see Searle (1993), Chalmers (1996), and Siewert (1998). Representational accounts of consciousness have often

been motivated by the thought that it is easier to see how intentional states could be reduced to physical states than how consciousness could be. In my view, which I do not argue for here, intentionality is ultimately to be understood as a form of consciousness, rather than the other way around, dispositional intentional states deriving their content from their manifestation in consciousness. If so, the question of the relation of consciousness to the physical is basic.

- 13 Importantly, I do not characterize the class of physical properties here as per se non-mental, though given the list of basic properties, they are clearly not mental per se. This leaves it open that mental properties could be analyzed as logical constructions of primary qualities, or, as conceptually supervening on them (see section 1.4).
- 14 See Poland (1994: esp. pp. 109–47) and Papineau (1993: 29–32).
- 15 More properly, a fully meaningful predicate in a language  $L$  expresses a concept and picks out a property. In different languages the same word may express different concepts, or none. I omit this relativization for brevity, but it should be understood as implicit wherever we are concerned with the relation of linguistic items to truth, concepts, and properties. I also ignore, for the most part, complications introduced by tense and other context-sensitive elements in natural languages.
- 16 There are other concepts of property that might be, and sometimes are, employed on which this would not be true. For example, one might individuate properties in terms of the sets of possible individuals who possessed them. Then two predicates would pick out the same property iff they were necessarily coextensive, which does not require synonymy (e.g., ‘is trilateral’ and ‘is triangular’). But the theses about property identity that could be expressed in this way can be expressed without the dubious ontology and unhelpful innovation in terminology, which should not be encouraged.
- 17 More generally, we would speak of sentences as analytic relative to occasions of utterance, since what many sentences express in natural languages is relative to context of utterance.
- 18 There is controversy about whether there are analytic statements, conceptual truths, and truths knowable a priori, but in stating the mind–body problem it is not necessary to take a stand on this. W. V. Quine’s “Two Dogmas of Empiricism” (1953) is the *locus classicus* of the case against analyticity. Grice and Strawson (1956) is an important early reply.
- 19 ‘Supervenience’ in its current use is usually said to have been introduced in the context of ethical theory by R. M. Hare in the early 1950s to describe the relation of ethical properties to natural properties, and then imported into discussions in the philosophy of mind by Davidson (1980). It was in use earlier in the emergentist tradition, though perhaps not with quite as specific a meaning; see Kim (1993b: essay 8).
- 20 There are many changes one can ring on this formulation. For example, if we put in ‘it is conceptually necessary that’ before the whole right-hand side of the biconditional, we get a version of what has been called strong supervenience (Kim 1993b: essay 4). There are weaker varieties as well. I use this formulation because I wish to allow conceptual supervenience of the mental on the physical even though there could be a world of non-material objects that had mental properties. This is a possibility which functionalism, for example, leaves open. This gives content to the idea that supervenience is strictly weaker than reduction. Sometimes supervenience claims are formulated in terms of indiscernibility claims:  $F$ -properties supervene on  $G$ -properties iff necessarily

things which are alike with respect to their G-properties are alike with respect to their F-properties. See the essays in Kim (1993b) and Savellos and Yalçin (1995) for further discussion of the variants and their relations to one another.

- 21 The requirement that psychological laws (including any psychophysical laws) be entailed by physical laws is needed to avoid the problem of lucky materialism (Witmer 2000).
- 22 This position may appear stronger than it is. I put no constraints on physical properties other than that they be physical. Complex relational properties may figure in the supervenience base. Thus, it is equivalent to the view that a complete physical description of the world entails a complete psychological account of it.
- 23 It has been used in a weaker sense to denote a materialist ontology, and in a stricter sense, e.g., by the Logical Positivists, to mean that all statements are translatable into the vocabulary of physics.
- 24 By ‘non-mental properties’ here I mean properties that are classified in terms that are not mental as such, so that some members of the class, and certainly all basic (i.e., non-complex) members, are not mental. This allows that mental properties may be a subclass of the properties in question. That is to say, (2) asserts that there are no classes of properties that are not mental per se to which mental properties are conceptually reducible.
- 25 In the present context, by a non-relational property we mean a property that an individual has which does not require the existence of some contingently existing individual not identical with the individual possessing the property or any part of it, and does not require the non-existence of any thing or kind of thing. For example, being married and being a planet are relational properties, being round and being red are not.
- 26 This rules out appeal to properties that constituents have because of emergent properties of the wholes they compose.
- 27 This leaves open that they may have mental properties in the sense that they have relational properties which entail that something possesses mental properties, e.g., because they coexist with or are part of a thing that has irreducible mental properties but which is not itself a basic constituent of things. Also this leaves open that the basic constituents of things have properties which we might not recognize as broadly physical, but it does not allow that they be mental. Thus, constituent non-mentalism is a more liberal thesis than constituent physicalism.
- 28 See Bealer (1992) for a general defense of these methods for discovering what is necessary and possible; a more recent book-length defense of conceptual analysis is Jackson (1998).
- 29 The discovery of this paradox by Bertrand Russell, in May 1901, played an important role in foundational studies in set theory and mathematics early in the twentieth century.
- 30 The question of the relation of consciousness and intentionality becomes important here, for the thought experiments mentioned seem to depend on our thinking that a conscious point of view could be missing in a being physically and behaviorally like us, or be present in a being with no associated body at all. If intentional states and conscious states are independent, the support of these thought experiments for the irreducibility of the mental *tout court* is reduced.
- 31 With the exception, however, of the role of the notion of an observation in quantum mechanics: how seriously this is to be taken is a matter of controversy.

- 32 See Annas (1992) for survey of ancient philosophy of mind concentrating on the Hellenistic period.
- 33 This was a minority position in antiquity. Introduced by Plato, it assumed the importance it has in the later western tradition through the influence of Plato’s philosophy on Catholic theology, through which it has permeated ideas about mind and body in western culture.
- 34 See Woolhouse (1993) for discussion of the notion of substance in early modern philosophy rationalists.
- 35 We must exclude here such “formal” properties as having a property.
- 36 The initial moves in the argument are made in the second meditation of Descartes’s masterpiece *Meditations on First Philosophy* (1985 [1641]) and concluded in the sixth; see also *Principles of Philosophy* (1985 [1644]: §63).
- 37 This analogy was conceived by Leibniz, though his basic metaphysics rejects substance dualism.
- 38 Though dualism is not currently a popular view among philosophers or scientists, it is still no doubt one of the most commonly, if unreflectively, held views about the relation of mental to physical phenomena, as it is the background metaphysics of a number of the world’s major religions; and it is not without contemporary proponents among philosophers and scientists, see, e.g., Foster (1996), Eccles (1953: ch. 8), Popper and Eccles (1977: ch. E7).
- 39 Three landmarks of the twentieth century are Carnap (1928), Lewis (1929), and Goodman (1951). A more recent proponent is Grayling (1985).
- 40 A detailed bibliography of sources is available at the end of the article on panpsychism in Edwards (1967).
- 41 Panpsychism, and other mental particle theories, as for reductive materialism, is an expression of the idea, as Popper and Eccles have put it, that there is nothing new under the sun, which is an expression of a form of the principle of sufficient reason: nothing can come from nothing (Popper and Eccles 1977: 14).
- 42 Nineteenth-century double aspect theorists include Shadworth Hodgson (1870: esp. ch. 3) and G. H. Lewes (1877).
- 43 See, e.g., the discussion of Morton Prince (1885; repr. in Vesey 1964: 187).
- 44 Perhaps Strawson’s view that the concept of a person is more basic than that of a person’s mind or body may be construed as of this sort (1958).
- 45 How should we classify a view such as Hume’s “bundle” theory of the self? On this view, there is no thing that is the self, but rather each self is to be construed as constituted out of a set of perceptions which bear appropriate relations to one another. The perceptions are intrinsically mental in character, like mental atoms. They are not, though, apparently thought of as in space. So, while a mental particle view of a sort, it is more like substance dualism without the basic mental substances being thinking beings, but rather thoughts constitutive of thinking beings. If we take “perceptions” to be non-mental themselves, and take both the self and ordinary objects to be logical constructions out of them, we arrive at a version of the neutral monism advocated by James, Mach, and Russell (see below).
- 46 It is important to distinguish emergence in this sense from emergence of higher levels of organization of complex systems governed by simple rules that is often discussed in the context of “chaos” theory. The properties of the latter sort conceptually supervene on the rules governing the constituents of the system, their properties, and their

arrangement. They are emergent only in the sense of being surprising to us, and so their status as emergent, in this sense, is a function of our inability to easily predict them.

- 47 This requires us to disallow indefinitely long disjunctions from expressing relevant types; otherwise, by disjoining all the nomically sufficient conditions stated in physical terms for a given mental type, we could always arrive at a nomically necessary condition.
- 48 A token is an instance of a type. For example, in the previous sentence (inscription), there are four tokens of the letter “a.” Tokens are always particulars. Every token is identical with itself. We get informative statements about token identity when we use different ways of picking out the same thing. It can be informative, e.g., to be told that Pluto is the smallest planet in the Solar System. Type-type identity, strictly speaking, is about properties. Again, every property is identical to itself and to no distinct thing. Informative type-type identity statements pick out the properties in different ways. We will see an example below of a type-type identity theory of the mental and the physical that makes this an interesting empirical discovery.
- 49 The conception of events articulated in note 4 is incompatible with anomalous monism, for it individuates events in terms of the objects and properties that they are changes with respect to. Thus, unless mental properties are physical properties, which on this view they are not, no mental event is token identical with any physical event. There are various weaker relations that could be articulated. For example, it might be said that every mental event occurs at the same time as and in the same object as a physical event. In any case, it is not clear that much hinges on this. The more fundamental question is about objects and properties rather than events.
- 50 In origin, a medical term meaning “symptom of an underlying cause” or “secondary symptom.”
- 51 See McLaughlin (1992) for a discussion of this particular school in the broader emergentist tradition. Be aware that McLaughlin uses ‘emergentism’ in a narrower sense than it is used here, namely, to cover what I would call emergent materialism with downward causation. ‘Emergentism’ is the right term for the rejection of (3); we can distinguish epiphenomenal and non-epiphenomenal versions, the latter of which will at least include emergentism with downward causation. Alas, terminological variation in philosophy is endemic. Broad himself, who introduced the term ‘emergent materialism’, did not take it to imply downward causation, which he accepted tentatively as an empirical hypothesis on the basis of what he took to be the evidence of psychical research.
- 52 This contrast and debate between epiphenomenal emergentists and downward causation emergentists reprises a similar debate in antiquity between followers of Aristotle (Caston 1997).
- 53 For more recent discussions, see Armstrong (1968: 47) and Kim (1993a,b).
- 54 A note is in order on the term ‘property dualism’, which has figured prominently in recent literature on the mind–body problem. This label is often used in application to emergentism, but applies to any position that holds that there are objects that have mental properties, and there are objects that have physical properties and that both sorts are basic properties, not conceptually reducible to each other or anything more basic. (Property dualism is not coextensive with any position that holds (1) and (2) and either of (3) or (4), since, e.g., idealism embraces (1)–(3), but reduces what are ordinarily thought of as physical properties to mental properties.) *Property* dualism

is a weaker view than *substance* dualism, but is entailed by it. “Property dualism” is often used as a term of abuse by philosophers attracted by reductionism, with the idea of associating its proponents with the discredited view of substance dualism by the overlap in the spelling of their labels. The introduction of “property dualism” into the philosophical vocabulary is not an entirely happy terminological innovation, and that is one reason it does not figure prominently in my discussion. Quite apart from its association with demagoguery, the label falsely suggests that there are at most two families of properties irreducible to each other: but even setting aside the current issue, there are many mutually irreducible families of properties (color and shape properties, for example).

- 55 There are some possible though unoccupied positions here that we will not survey, such as the view that the mental supervenes on or is conceptually reducible to something non-mental, and the physical in turn supervenes on or is conceptually reducible to the mental.
- 56 In this way it is like the reduction of mathematics to logic and set theory. We can retain our old forms of speech, but our ontology includes only sets, not numbers in addition. The relation of the mental and physical to underlying reality on neutral monism is like the relation of odd and even numbers to the underlying reality on the set-theoretic reduction of mathematics. Each is distinct from the other, and has an essential property the other cannot have, but each is explained as a logical construction out of something more basic.
- 57 Carnap (1931) and Hempel (1935) provide early examples of logical behaviorists; both later retreated from the early position. Ryle’s *The Concept of Mind* (1949) was an important and influential behaviorist manifesto (though Ryle denied the term applied to his view). Wittgenstein’s *Philosophical Investigations* (1950) was an important inspiration for criterial behaviorism. See, for example, Malcolm (1958). Important psychological behaviorists were Watson (1925) and Skinner (1974), though their behaviorism was methodological rather than logical.
- 58 See Putnam (1968). Logical behaviorism seems to have succumbed to a danger that every reductive project faces. As C. I. Lewis put it: “Confronted with problems of analysis which there is trouble to resolve, one may sometimes circumvent them by changing the subject” (1941: 225).
- 59 This is held to be true as a matter of fact. Of course, if there were non-physical objects that had internal structure, they would have functional states as well.
- 60 Every era has its favored metaphor provided by its prestige technology. In the seventeenth and eighteenth centuries, it was the clock or the mill. In the nineteenth, it was the steam engine. In the latter half of the twentieth, it became the computer.
- 61 The Pythagoreans advocated the general idea of functionalism, that having a mind depends on a certain organization of the body, in antiquity. It is one of the positions that Socrates responds to in the *Phaedo* in Simmias’s suggestion that the soul is to the body as the attunement of it is to a string instrument (Plato 1989: 69).
- 62 There are two ways of understanding psychofunctionalism’s empirical character. First, it can be understood as a version of emergentism with bridge laws connecting functional with mental states (see, e.g., Chalmers 1996: ch. 6). Second, it might be maintained that the identification of mental with functional states is a theoretical identification, like the identification of lightning as an electrical discharge (this view has been advocated for intentional states in Rey 1997). If what I say below about this

is correct, this introduces an empirical element into the discovery, but not in a way that prevents this view, if correct, from collapsing into analytic functionalism. See the discussion of the identity theory below.

- 63 A more recent variant on the general theme is connectionism. A connectionist system consists of a set of interconnected units that can take on activation values: the interconnections determine the influence of the activation value of a given node on those connected to it. Through their connections, units may inhibit or excite other units to various degrees depending on their own activation states. Certain units may be designated input units and others output units. The activation values can be continuous, so a connectionist system is not a finite state machine. But it fits our initial very general characterization of a functional system, since different connectionist systems are wholly characterized in terms of their states' relations to input and output and other states. The difference between classical functionalism and connectionism will not be relevant at the level of our discussion here.
- 64 The force of this objection is unclear. Either mental properties are analyzable as functional properties or not. If not, then there is the question whether they are analyzable as physical properties. If so, that is an end to the matter, and the charge of chauvinism is bootless. If not, it is an empirical matter what physical state types, if any, mental state types are correlated with, and our hunches or prejudices about it are irrelevant. Though in the latter case, clearly difficulties will arise when we try to confirm or disconfirm claims about physical systems that are very different from ourselves.
- 65 See also Searle's Chinese Room thought experiment (1980), and Chalmers (1996: ch. 3) for a recent deployment of so-called zombie thought experiments to establish the irreducibility of conscious mental states.
- 66 Putnam was careful to exclude systems that contain parts that have organizations like the whole they constitute. This would rule out the system in Block's thought experiment as constituting a person. However, it is difficult to see what justifies the exclusion. For if our mental concepts are functional concepts, it should not matter how the system that has the appropriate functional organization is constituted.
- 67 See Chalmers (1996: 91–101) for a somewhat fuller discussion and some responses to objections that have appeared in the literature.
- 68 A functionalist need not require this. A functional system could be characterized in terms of non-causal transitions between states given input and output. But this opens the door to a great many more machine table descriptions of objects that may have minds than a functionalist will typically want to countenance.
- 69 The event reported in the headlines of this morning's paper caused extensive flooding in coastal areas of Florida, but it was not by virtue of being of that type that it did so, but in virtue of its being the passing of a category 3 hurricane off the coast. Causal relations hold between particulars, datable events, or states. But to explain why they hold between those particulars we must appeal to their types.
- 70 See Jackson and Pettit (1988), Block (1990), Fodor (1991), Dardis (1993), and Ludwig (1994a, 1998) for discussion.
- 71 At this point too the question whether intentional states are conceptually independent of conscious states is important, for our conviction that mental states are causally relevant to behavior seems to attach in the first instance to conscious mental states, and to dispositional states only through their manifestation conditions in consciousness.

For dispositional states too are defined in terms of manifestation conditions, and so as such are not causally relevant to those conditions.

- 72 The account given here departs from Lewis's own. Perhaps the departure is largely in terminology, but it is still important. Lewis has argued that despite the theoretical identification of pain with a physical property in human beings, it still makes sense to say that some being (a Martian, e.g.) could be in pain though he does not have that property which in human beings is pain (Lewis 1980). How is this possible? It is not, if we understand the relation between predicates, concepts, and properties as I have introduced them. On the account I have given, the predicate 'is in pain' expresses the concept of pain and is used to attribute the property of being in pain, and it does each of these in virtue of its meaning in English. The property is, so to speak, the shadow of the meaning of the predicate cast on the world, and the concept is the shadow it casts in our thoughts. If the property of being in pain is a physical property, so, on this view, is the concept of pain a physical concept. Lewis, however, identifies something else as the concept of pain. To put it briefly, Lewis uses 'concept of pain' to denote the concept expressed by the predicate 'is a thing that has the property  $P$  such that, for the most part,  $T(P)$  for beings of kind  $K$ ' where ' $T(P)$ ' is replaced by the appropriate psychological theory with ' $P$ ' in the place of the variable representing the property of pain. That concept applies to a thing in virtue of that thing's having some property that plays a certain role mediating input and output. It might have been that a different property played that role. And in different kinds of beings, perhaps, for the most part, different properties play that role. However, Lewis does not say that the property of being in pain is the one attributed using this form of predicate. Rather, Lewis calls the property that actually plays the role the property of being in pain. This allows then that in different kinds of beings a different property can be (called) the property of being in pain. It also apparently allows that if things had been different, a different property in us would have been (called by us) the property of being in pain. Apparently, however, Lewis does want to treat the predicate 'is in pain' as if it attributed the property that plays the right role. Thus, he says "is in pain" is ambiguous when we apply it to different kinds of beings, and when we consider it in different possible worlds. For a difference in property attributed entails a difference in the meaning of the predicate. It is as if we had decided to say that the property being rich is attributed using 'has a lot more money than most people' but the concept of richness is expressed by the predicate 'is Ludwig's favorite property'. I keep the concept of pain attached to the predicate 'is in pain', and so matched with the property attributed using it. This follows the traditional alignment, and provides us a clearer view of the issues.
- 73 There are many arguments against the psychophysical identity theory and physicalism more generally that rest on thought experiments designed to show that nothing follows about what mental properties an object has from an exhaustive description of its physical construction. One style of argument much discussed recently has been dubbed 'the knowledge argument'. Some deployments of the argument in the latter half of the twentieth century are Meehl (1966), Nagel (1979b), Jackson (1982, 1986). Leibniz already gives a version of such an argument in *The Monadology* (1714: sec. 16): "If we imagine that there is a machine whose structure makes it think, sense, and have perceptions, we could conceive it enlarged, keeping the same proportions, so that we could enter into it, as one enters into a mill. Assuming that, when inspecting

its interior, we will only find parts that push one another, and we will never find anything to explain a perception.” These arguments are certainly decisive against any version of the psychophysical identity that suggests that we can perform an armchair analysis of our mental concepts to determine that they in fact pick out neurophysiological properties. They do not address versions of the theory that treat our ordinary terms as having their concepts fixed by description: the burden such an approach takes up, though, is scarcely less heavy, for it must allow, as we have seen, that our terms may fail to express any concepts at all.

- 74 The view itself was certainly not undiscussed previously in the twentieth century. Broad discusses and dismisses it (1925: 622–3). C. I. Lewis discussed and criticized a form of the identity theory, which he presents as proposing descriptive definitions of mental terms, in much the same spirit as the theory I have presented (1941: 230–1). Some of Smart’s replies to objections are clearly directed at Broad’s and Lewis’s earlier discussions.
- 75 They regarded propositional attitudes as understandable behavioristically, or functionally. However, the position can easily be generalized to propositional attitudes.
- 76 See Bealer (1987, 1994) for arguments against this appeal to what is sometimes called scientific essentialism.
- 77 In any case, it should be noted that the same unclarity would attach to whatever notion of property identity would be here invoked as attaches to metaphysical necessity: if we try to explain it in accordance with the tradition, we must admit that what we discover is that, e.g., “water” and “H<sub>2</sub>O” express the same concept, contrary to the supposition.
- 78 These began with work by Kripke (1980) on proper names and natural kind terms and Hilary Putnam (1975) on natural kind terms in the early 1970s. Initially, these arguments were directed toward showing that the meanings of various natural language terms were determined by their causal relations with things and kinds in our environments. Since we use these same terms to characterize our attitudes, however, it was soon apparent that these arguments might be used to urge also that our thought contents were individuated relative to what things and kinds were actually in our environments.
- 79 See Putnam (1975) and Burge (1979, 1982, 1986). Widespread uncritical acceptance of externalism is a salient feature of discussion in contemporary philosophy of mind.
- 80 Difficulties are discussed in Ludwig (1992a, 1992b, 1993a, 1993b, 1993c, 1994a, 1994b, 1996a, 1996b).
- 81 The literature on this subject is large. An earlier paper that advanced this thesis particularly in response to Putnam (1981) is Brueckner (1986). See also Boghossian (1989, 1993).
- 82 See Jackson (1996) for a fairly comprehensive review of discussion of mental causation.
- 83 Early proponents were Feyerabend (1963) and Rorty (1965, 1979). Perhaps Wittgenstein endorsed eliminativism in the *Tractatus Logico-Philosophicus* (1921), but if so on grounds more abstract than more recent eliminativists. Eliminativism may be the one modern view that is not represented in ancient philosophy. Perhaps the atomists, Leucippus and Democritus, might be thought to endorse eliminativism, since they held that reality consisted solely of atoms and the void. But they showed no inclination to deny that there were people who thought and reasoned, and

Democritus seems to have intended to explain psychological phenomena in terms of his atomistic metaphysics (see Taylor 1999). One can be a partial, as well as a wholesale eliminativist. Georges Rey argues for functionalism for intentional states, but eliminativism for qualitative states (1997).

### References

- Alexander, S. (1920). *Space, Time and Deity: The Gifford Lectures at Glasgow 1916–1918*. London: Macmillan.
- Anderson, A. R. (ed.) (1964). *Minds and Machines*. Englewood Cliffs: Prentice Hall.
- Annas, J. (1992). *Hellenistic Philosophy of Mind*. Berkeley: University of California.
- Aristotle. (1984). *The Complete Works of Aristotle*. Princeton: Princeton University Press.
- Armstrong, D. M. (1968). *A Materialist Theory of the Mind*. London: Routledge and Kegan Paul, Humanities Press.
- Ayer, A. J. (ed.) (1959). *Logical Positivism*. New York: Free Press.
- Beakley, B. and Ludlow, P. (eds.) (1992). *The Philosophy of Mind: Classical Problems/Contemporary Issues*. Cambridge, MA: MIT Press.
- Bealer, G. (1987). “The Philosophical Limits of Scientific Essentialism.” *Philosophical Perspectives*, 1: 289–365.
- (1992). “The Incoherence of Empiricism.” *Proceedings of the Aristotelian Society*, 66: 99–138.
- (1994). “Mental Properties.” *The Journal of Philosophy*, 91: 185–208.
- Beckermann, A., Flohr, H. et al. (1992). *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*. New York: W. de Gruyter.
- Bedau, M. (1986). “Cartesian Interaction.” *Midwest Studies in Philosophy*, 10: 483–502.
- Berkeley, G. (1710). *A Treatise Concerning the Principles of Human Knowledge*. In Michael R. Ayers (ed.), *Philosophical Works*. London: The Guernsey Press, 1975: 61–128.
- Block, N. (1978). “Troubles with Functionalism.” In C. W. Savage (ed.), *Perception and Cognition: Issues in the Foundations of Psychology*. Minneapolis: University of Minnesota: 261–325. Reprinted in Block (1980).
- (ed.) (1980). *Readings in the Philosophy of Psychology*. Cambridge, MA: Harvard University Press.
- (1990). “Can the Mind Change the World?” In G. Boolos (ed.), *Meaning and Method: Essays in Honor of Hilary Putnam*. New York: Cambridge University Press: 29–59.
- Block, N., Flanagan, O. et al. (eds.) (1997). *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA: MIT Press.
- Boghossian, P. (1989). “Content and Self-Knowledge.” *Philosophical Topics*, 17: 5–26.
- (1993). “The Transparency of Content.” *Philosophical Perspectives*, 8: 33–50.
- Borst, C. V. (ed.) (1970). *The Mind–Brain Identity Theory*. London: Macmillan.
- Brentano, F. (1955 [1874]). *Psychologie vom Empirischen Standpunkt (Psychology from an Empirical Standpoint)*. Hamburg: Felix Meiner.
- Broad, C. D. (1925). *The Mind and Its Place in Nature*. New York: Harcourt, Brace and company.
- Brueckner, A. (1986). “Brains in a Vat.” *The Journal of Philosophy*, 83: 148–67.
- Burge, T. (1979). “Individualism and the Mental.” *Midwest Studies in Philosophy*, 4: 73–121.

- (1982). “Other Bodies.” In A. Woodfield (ed.), *Thought and Object: Essays on Intentionality*. Oxford: Clarendon Press: 97–120.
- (1986). “Cartesian Error and the Objectivity of Perception.” In P. Pettit (ed.), *Subject, Thought, and Context*. New York: Clarendon Press: 117–36.
- Carnap, R. (1928). *The Logical Structure of the World*, trans. Rolf A. George. Berkeley: University of California Press, 1967. First published in German in 1928 under the title *Der Logische Aufbau der Welt*.
- (1931). “Die Physikalische Sprache als Universalsprache der Wissenschaft.” *Erkenntnis*, 2: 432–65. Reprinted in Ayer (1959) as “Psychology in Physical Language.”
- Caston, V. (1997). “Epiphenomenalisms, Ancient and Modern.” *The Philosophical Review*, 106: 309–64.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Churchland, P. (1981). “Eliminative Materialism and the Propositional Attitudes.” *The Journal of Philosophy*, 78: 67–90.
- Cooney, B. (ed.) (2000). *The Place of Mind*. Stamford: Wadsworth.
- Dardis, A. (1993). “Sunburn: Independence Conditions on Causal Relevance.” *Philosophy and Phenomenological Research*, 53: 577–98.
- Davidson, D. (1980). ‘Mental Events’. In *Essays on Actions and Events*. New York: Clarendon Press: 207–24.
- Descartes, R. (1985 [1641]). *Meditations on First Philosophy*. In *The Philosophical Writings of Descartes*, vol. II, ed. and trans. J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Cambridge University Press: 1–62.
- (1984 [1644]). *The Principles of Philosophy*. In *The Philosophical Writings of Descartes*, vol. I, ed. and trans. J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Cambridge University Press: 177–292.
- Dretske, F. (1997). *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- Eccles, J. C. (1953). *The Neurophysiological Basis of Mind; The Principles of Neurophysiology*. Oxford: Clarendon Press.
- (ed.) (1985). *Mind and Brain: The Many-Faceted Problems*. New York: Paragon House.
- Edwards, P. (1967). *The Encyclopedia of Philosophy*. New York: Macmillan.
- Feigl, H. (1958). “The Mental and the Physical.” In H. Feigl et al., *Concepts, Theories and the Mind–Body Problem*. Minneapolis: University of Minnesota Press: 370–497.
- Feyerabend, P. (1963). “Mental Events and the Brain.” *The Journal of Philosophy*, 60: 295–96.
- Fodor, J. A. (1991). “A Modal Argument for Narrow Content.” *The Journal of Philosophy*, 88 (1): 5–26.
- Foster, J. (1996). *The Immaterial Self: A Defence of the Cartesian Dualist Conception of the Mind*. London: Routledge.
- Grayling, A. C. (1985). *The Refutation of Scepticism*. LaSalle, IL: Open Court.
- Grice, H. P. and Strawson, P. F. (1956). “In Defense of a Dogma.” *Philosophical Review*, 65: 141–58.
- Goodman, N. (1951). *The Structure of Appearance*. Cambridge, MA: Harvard University Press.
- Hempel, C. (1935). “The Logical Analysis of Behavior.” In Block (1980: 14–23). Originally published in *Revue de Synthèse*, 1935.

- Hodgson, S. (1870). *The Theory of Practice*. London: Longmans, Green, Reader and Dyer.
- Huxley, T. H. (1874). “On the Hypothesis that Animals are Automata, and its History.” In Vesey (1964: 134–43). Originally delivered in 1874 to the British Association for the Advancement of Science, Belfast.
- (1901). *Methods and Results: Essays*. New York: D. Appleton.
- Jackson, F. (1982). “Epiphenomenal Qualia.” *The Philosophical Quarterly*, 32: 127–36.
- (1986). “What Mary Didn’t Know.” *The Journal of Philosophy*, 83: 291–5.
- (1996). “Mental Causation.” *Mind*, 105 (419): 377–413.
- (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. New York: Clarendon Press.
- Jackson, F. and Pettit, P. (1988). “Functionalism and Broad Content.” *Mind*, 97: 381–400.
- James, W. (1950 [1890]). *The Principles of Psychology*. New York: Dover.
- (1910). *Psychology*. New York: Henry Holt and Co. Page references in the text are given to Block et al. (1997), in which a portion of chapter 11 is reprinted.
- (1904). “Does Consciousness Exist?” *The Journal of Philosophy, Psychology and Scientific Methods*, 1: 477–91. Repr. in James (1976 [1912]).
- (1976 [1912]). *Essays in Radical Empiricism*. Cambridge, MA: Harvard University Press.
- Kant, I. (1997 [1781]). *The Critique of Pure Reason*. Cambridge: Cambridge University Press. Originally published in 1781, a much revised version was published in 1787.
- Kim, J. (1993a). “Mechanism, Purpose, and Explanatory Exclusion.” In Kim (1993b: 237–64).
- (1993b). *Supervenience and Mind*. New York: Cambridge University Press.
- Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Leibniz, G. W. (1714). *Principles of Philosophy, or, The Monadology*. In *Philosophical Essays*. Indianapolis: Hackett, 1989: 213–24. Originally composed in 1714 for private correspondence as a summary of his philosophical views.
- Lewes, G. H. (1877). *The Physical Basis of Mind*. London: Trubner and Co.
- Lewis, C. I. (1929). *Mind and the World-order: Outline of a Theory of Knowledge*. New York: Scribner.
- (1941). “Some Logical Considerations Concerning the Mental.” *The Journal of Philosophy*, 38: 225–33.
- Lewis, D. (1966). “An Argument for the Identity Theory.” *The Journal of Philosophy*, 63: 17–25.
- (1972). “Psychophysical and Theoretical Identifications.” *The Australasian Journal of Philosophy*, 50: 249–58. Repr. in Lewis (1999).
- (1980). “Mad Pain and Martian Pain.” In Block (1980: 216–22).
- (1999). *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
- Ludwig, K. (1992a). “Skepticism and Interpretation.” *Philosophy and Phenomenological Research*, 52 (2): 317–39.
- (1992b). “Brains in a Vat, Subjectivity, and the Causal Theory of Reference.” *Journal of Philosophical Research*, 17: 313–45.
- (1993a). “Direct Reference in Thought and Speech.” *Communication and Cognition*, 26 (1), 49–76.
- (1993b). “Externalism, Naturalism, and Method.” In E. Villanueva (ed.), *Naturalism and Normativity*. Atascadero, Ridgeview: 250–64.

- (1993c). “Dretske on Explaining Behavior.” *Acta Analytica*, 811: 111–24.
- (1994a). “Causal Relevance and Thought Content.” *Philosophical Quarterly*, 44: 334–53.
- (1994b). “First Person Knowledge and Authority.” In G. Preyer, F. Siebelt, and A. Ulfig (eds.), *Language, Mind, and Epistemology: On Donald Davidson’s Philosophy*. Dordrecht: Kluwer: 367–98.
- (1996a). “Singular Thought and the Cartesian Theory of Mind.” *Nous*, 30 (4): 434–60.
- (1996b). “Duplicating Thoughts.” *Mind and Language*, 11 (1): 92–102.
- (1996c). “Explaining Why Things Look the Way They Do.” In K. Akins (ed.), *Perception*. Oxford: Oxford University Press: 18–60.
- (1998). “Functionalism and Causal Relevance.” *Psyche* 4 (3) (<http://psyche.cs.monash.edu.au/v4/psyche-4-03-ludwig.html>).
- Lycan, W. G. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.
- (ed.) (1990). *Mind and Cognition: A Reader*. Oxford: Blackwell.
- Mach, E. (1897 [1886]). *Ernst Mach: Contributions to the Analysis of Sensations*. Chicago: Open Court.
- Malcolm, N. (1958). “Knowledge of Other Minds.” *The Journal of Philosophy*, 55: 969–78. Reprinted in Rosenthal (1991).
- McGinn, C. (1989). “Can We Solve the Mind–Body Problem?” *Mind*, 98: 349–66.
- (1991). *The Problem of Consciousness*. Oxford: Blackwell.
- (1999). *The Mysterious Flame: Conscious Minds in a Material World*. New York: Basic Books.
- McLaughlin, B. (1992). “The Rise and Fall of British Emergentism.” In Beckermann (1992).
- Meehl, P. E. (1966). “The Compleat Autocerebroscopist: A Thought-Experiment on Professor Feigl’s Mind–Body Identity Thesis.” In P. K. Feyerabend and G. Maxwell (eds.), *Mind, Matter, and Method: Essays in Philosophy and Science in Honor of Herbert Feigl*. Minneapolis: University of Minnesota Press: 103–80.
- Menzies, P. (1988). “Against Causal Reductionism.” *Mind*, 97: 551–74.
- Moore, G. E. (1959 [1903]). “The Refutation of Idealism.” *Philosophical Studies*. Paterson, N.J.: Littlefield, Adams and Co.: 1–30. Originally published in *Mind*, 13: 433–53.
- Morgan, C. L. (1923). *Emergent Evolution: The Gifford Lectures*. London: Williams and Norgate.
- Nagel, T. (1979a). “Panpsychism.” In *Mortal Questions*. Cambridge: Cambridge University Press: 181–95.
- (1979b). “What Is It Like to Be a Bat?” In *Mortal Questions*. Cambridge: Cambridge University Press: 165–80.
- (1994). “Consciousness and Objective Reality.” In Richard Warner (ed.), *The Mind–Body Problem: A Guide to the Current Debate*. Oxford: Blackwell: 63–8.
- (1998). “Conceiving the Impossible and the Mind–Body Problem.” *Philosophy*, 73 (285): 337–52.
- O’Connor, J. (ed.) (1969). *Modern Materialism: Readings on Mind–Body Identity*. New York: Harcourt, Brace and World.
- Papineau, D. (1998). “Mind the Gap.” *Philosophical Perspectives*, 12: 373–88.
- (1993). *Philosophical Naturalism*. Oxford: Blackwell.
- Place, U. T. (1956). “Is Consciousness a Brain Process?” *British Journal of Psychology*, 45: 243–55.

- Plato (1989). *The Collected Dialogues of Plato*. Princeton: Princeton University Press.
- Poland, J. S. (1994). *Physicalism, the Philosophical Foundations*. Oxford: Clarendon Press.
- Popper, K. (1994). *Knowledge and the Body–Mind Problem: In Defence of Interaction*. London: Routledge.
- Popper, K., and Eccles, J. (1977). *The Self and its Brain*. New York: Springer-Verlag.
- Prince, M. (1885). *The Nature of Mind and Human Automatism*. Philadelphia: J. B. Lippincott Company.
- Putnam, H. (1967). “Psychological Predicates.” In W. H. Capitan and D. D. Merrill (eds.), *Art, Mind and Religion*. Pittsburgh: Pittsburgh University Press: 37–48. Reprinted under the title “The Nature of Mental States” in Rosenthal (1991).
- (1968). “Brains and Behavior.” In J. R. Butter (ed.), *Analytic Philosophy*. Oxford: Basil Blackwell: 1–19. Reprinted in Rosenthal (1991).
- (1975). “The Meaning of ‘Meaning’.” In *Mind, Language and Reality: Philosophical Papers*. Cambridge: Cambridge University Press: 215–71.
- (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Quine, W. V. O. (1953). “Two Dogmas of Empiricism.” In *From a Logical Point of View*. Cambridge, MA: Harvard University Press: 20–46.
- Rey, G. (1997). *Contemporary Philosophy of Mind: A Contentiously Classical Approach*. New York: Blackwell.
- Romanes, G. J. (1895). *Mind and Motion and Monism*. New York: Longmans Green.
- Rorty, R. (1965). “Mind–Body Identity, Privacy, and Categories.” *The Review of Metaphysics*, 19: 24–54.
- (1979). *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- Rosenthal, D. (ed.) (1971). *Materialism and the Mind–Body Problem*. Englewood Cliffs: Prentice-Hall.
- (1991). *The Nature of Mental States*. New York: Oxford University Press.
- Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson and Company, Ltd.
- Russell, B. (1917). “On the Relation of Sense-data to Physics.” In *Mysticism and Logic, and Other Essays*. London: Allen and Unwin: 108–31.
- (1921). *The Analysis of Mind*. London: G. Allen and Unwin.
- Savellos, E. and Yalçın, U. (eds.) (1995). *Supervenience: New Essays*. New York: Cambridge University Press.
- Searle, J. R. (1980). “Minds, Brains, and Programs.” *The Behavioral and Brain Sciences*, 3: 417–24. Reprinted in Rosenthal (1991).
- (1983). *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- (1984). “Intentionality and Its Place in Nature.” *Dialectica*, 38: 87–100.
- (1993). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- Siewert, C. (1998). *The Significance of Consciousness*. Princeton: Princeton University Press.
- Skinner, B. F. (1974). *About Behaviorism*. New York: Knopf.
- Smart, J. J. C. (1959). “Sensations and Brain Processes.” *The Philosophical Review*, 68: 141–56.
- Sperry, R. W. (1986). “Discussion: Macro- Versus Micro-Determination.” *Philosophy of Science*, 53: 265–70.
- Spinoza, B. (1994). *A Spinoza Reader: The Ethics and Other Works*. Princeton: Princeton University Press.

- Stich, S. (1983). *From Folk Psychology to Cognitive Science: The Case Against Belief*. Cambridge, MA: MIT Press.
- Strawson, P. F. (1958). "Persons." In Herbert Feigl, Michael Scriven, and Grover Maxwell (eds.), *Minnesota Studies in the Philosophy of Science 2*. Minneapolis: University of Minnesota Press: 330–53.
- Taylor, C. C. W. (1999). "The Atomists." A. A. Long (ed.), *The Cambridge Companion to Early Greek Philosophy*. Cambridge: Cambridge University Press: 181–204.
- Tye, M. (1997). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
- Vesey, G. N. A. (ed.) (1964). *Body and Mind: Readings in Philosophy*. London: George Allen and Unwin.
- Warner, R. and Szubka, T. (eds.) (1994). *The Mind–Body Problem: A Guide to the Current Debate*. Oxford: Basil Blackwell.
- Watson, J. B. (1925). *Behaviorism*. New York: W. W. Norton.
- Wilson, E. O. (1998). *Consilience: The Unity of Knowledge* (1st edn). New York: Knopf.
- Witmer, D. G. (2000). "Sufficiency Claims and Physicalism: A Formulation." In C. Gillett and B. Loewer (eds.), *Physicalism and Its Discontents*. New York: Cambridge University Press: 57–73.
- Wittgenstein, L. (1961 [1921]). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul Ltd. First published in 1921 in *Annalen der Naturphilosophie*.
- (1950). *Philosophical Investigations*. London: Macmillan.
- Woolhouse, R. S. (1993). *Descartes, Spinoza, Leibniz: The Concept of Substance in Seventeenth Century Metaphysics*. London: Routledge.