# A Logician's Perspective on the Relation Between the Mind and Body

# A Logician's Perspective on the Relation Between the Mind and Body:

## *Mentality and Modality*

By

John P. Burgess

Cambridge Scholars Publishing

A Logician's Perspective on the Relation Between the Mind and Body: Mentality and Modality

By John P. Burgess

This book first published 2024. The present binding first published 2024.

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library



ISBN (10): 1-0364-0189-8
ISBN (13): 978-1-0364-0189-4

for Ruby and Stella

“Material things I think, or feel, or see;
All else is immaterial to me.”
—Ambrose Bierce, *Devil’s Dictionary*

# CONTENTS

# PREFACE

This short book is on a topic familiar to all students of philosophy, and a subtopic familiar at least to specialists in the philosophy of mind, but it is written from an uncommon point of view. The topic is the mind-body problem, now almost four centuries old, going back as it does to René Descartes and Cartesian dualism. The subtopic is the thesis of the supervenience of the mental on the physical, under discussion for about half a century now since it was first adumbrated by Donald Davidson to be later elaborated by Jaegwon Kim. The point of view from which these topics are approached is that of an outsider come to the issues from modal logic rather than philosophical psychology.

The main conclusions to be suggested or insinuated in this work are two: first, that the supervenience thesis is false; but second, that whether it is true or false is nowhere near as important as has often been assumed. If these conclusions were accepted, one side in a long- and hard-fought contest would first be awarded the victory trophy, then told it is not gold but gilt. It would be unrealistic to hope to settle issues so long in dispute as those addressed here in so slim a volume as this, but I hope with the contrasting pair of conclusions just enunciated I may provide an irritant stimulus and do something to shake up the current state of debate.

In the text I have generally written as if addressing a newcomer, not previously acquainted with the questions under discussion. Why such a *faux naïf* approach? It comes down to this, that in looking at the literature of the subject from my outsider's perspective, I found myself on quite a few points inclined to doubt what seemed to be widely assumed among insiders. (I was shocked, for instance, at how many seem confident that it makes sense to attribute propositional attitudes to zombies.) I thus implicitly have to ask for a good deal in the way of suspension of disbelief from readers who might come to this book after having steeped themselves in the mainstream literature and absorbed its background assumptions.

Perhaps the only practicable route to the required suspension is through projecting oneself imaginatively into the situation of one encountering the issues for the first time. It is to assist such an effort of imagination that in the text I do not from the beginning assume, as I have in this preface, that the reader has already heard of dualism, supervenience, zombies, and so forth. In the text the first place where a word or phrase appears in a special

or technical sense it appears in **boldface**.

This work is in part a sequel to my *Saul Kripke: Puzzles and Mysteries* (2013) and the concluding chapter incorporates some related material from my Kripke Lecture at the CUNY Graduate Center in 2012, though no familiarity with my earlier work is presupposed.

My general intellectual debts remain those of the earlier book, now joined by the influence of key works of Ned Block, David Chalmers, Frank Jackson, Thomas Nagel, and (more in privately-circulated than published items) Stevan Harnad. I am grateful to Oliver Marshall, Neil Tennant, Alexander Williams, and Alexi Burgess for opportunities to present portions of this material publicly as it developed. I am grateful to the hosts and audiences at those public presentations, and the five leaders in the field just listed, and my colleagues Daniel Garber and Mark Johnston and former student Jack Woods, not to mention anonymous referees, for comments at greater or lesser length on earlier versions of parts of this work. Let me add a mention of Adriana Renero, though I was too far along in the production process to be able to make as much use as they deserve of her important study of Kripke (Renero, 2023) and her related remarks directly on an earlier draft of this work.

It was a cryptic remark of my late colleague David Lewis, to the effect that virtually *all* present-day positions in philosophy of mind would have been considered by Victorians to be forms of materialism, that first inspired me to undertake this study.

Princeton
January, 2024

# PART A.

# GHOSTS AND ZOMBIES

# CHAPTER 1

# DUALISM

## Dualism: Cartesian and Contemporary

Fifty years or so ago a number-theorist writing about the problem of representing an *n*th power as a sum of two *n*th powers and a philosopher of mind writing about the problem of the relation of the mental to the physical might well have expressed themselves in similar terms. "Work on the problem," such a commentator might have said, "constitutes only a restricted and rather specialized part of a large and growing field. But given the venerable age of the question, first brought to the attention of the learned world by a seventeenth-century French thinker," and here the allusion would be to Pierre de Fermat in the mathematical case, René Descartes in the philosophical case, "it will always have a certain romantic appeal, drawing to itself some of the most acute thinkers in each generation—and inevitably also any number of cranks." A writer addressing the philosophical problem might do so in similar terms today; by contrast, the mathematical problem is now solved, and the prize long on offer for its solution collected. Philosophical problems are more refractory than mathematical problems, some might conclude.

Virtually any expository writer addressing the **mind-body problem** would mention Descartes, as I have just done. Many would make a point of mentioning also what have been perceived as similar or related problems in other intellectual and cultural traditions. But it is to Descartes and a handful of his contemporaries that we primarily owe the two features most distinctive of *our* mind-body problem: first, the neural orientation, emphasizing the nervous system, especially the brain, and not the heart or liver or spleen or kidneys, as the part of the body most relevant to mental life; second, the mathematization of the concept of the material or physical that is contrasted with the mental or spiritual, arising from the seventeenth-century refocusing of science on what is quantifiable to the exclusion of what is not, for instance, on measurable size and shape, to the exclusion of sensed color. These two features are so much taken for granted today that there is a danger of forgetting that they once were contentious novelties.

The main topic of this study will be a third feature that the contemporary mind-body problem shares with Descartes' of four hundred years ago, namely, the curious entanglement of the issue mentality and physicality with the notions of possibility and necessity—or with **modality**, to give the category to which such notions belong its usual name. I will in the end plead for disentanglement, but to make a case for disentanglement is only my final destination, and I attach less importance to arriving at it than to journeying towards it, enlarging our understanding of mentality and modality along the way.

Now while a writer on the mind-body problem in a handbook or encyclopedia will doubtless mention Descartes, if working under any halfway serious word limit such a writer will be unlikely to linger long over the historical background. The writer may take the opportunity while briefly discussing such early modern figures as Berkeley and Hobbes and Descartes to introduce the labels that were retrospectively attached to their views: **idealism**, which makes everything mental; **materialism**, which makes everything physical; and **dualism**, which recognizes both mental and physical and so is left with the problem of what unites them. For dualism and materialism, at least, have substantial numbers of avowed present-day adherents, even if idealism does not. (I personally have never met more than two professed idealists.) But the writer will surely hasten to add that seventeenth- or eighteenth-century figures did not discuss the issues in the same terms in which one would today.

In jargon, it may be said that while Descartes advocated a "**substance dualism**," today what is more commonly met with is a "**property dualism**." What may be meant is that when self-identified dualists today speak of human beings—to which discussion will be limited here, with only occasional mention of the real or imagined mental lives of animals and extraterrestrials and robots—the emphasis is less on claiming that each consists of two irreducibly distinct components, a human mind and a human body, than on claiming that the life of each has two irreducibly distinct kinds of features, mental and physical. Talk of "**mind**" or "**soul**" may be little more than a way of talking collectively about the aggregate of mental features. What more might be meant will gradually unfold in what follows.

As for "property dualism," what I have vaguely called mental "features" are in the literature often called "properties"—but not always. Some parts of the literature are concerned instead with "mental events," other parts with "mental facts," and so on. Here I will most often write of "mental states," though sometimes adopting the chameleon-like policy of using whatever term is most used in the part of the literature I am discussing.

How widespread are dualist views among philosophers today? No general census has ever been taken. The only estimate I know of that is based on an actual survey, (Bourget and Chalmers, 2020) shows that among those polled about half accepted or leaned towards "**physicalism**," a third "non-physicalism," the remaining sixth various forms of "none of the above." Dualism, being a form of non-physicalism, can fairly confidently be said to be a minority view; anything beyond that would be a mere impression.

Now to bring out the issues about modality that are my ultimate official topic, there is no need to consider the whole range of views that are out there, but just a few samples, including one of a contemporary dualist type; and even it need not be specified in every respect, since not all aspects will be equally relevant to modal issues. The key features of the particular kind of contemporary dualism on which I wish to focus (henceforth called simply "contemporary dualism") are perhaps best introduced by comparisons and contrasts with Cartesian views. So let us stay a little bit longer than a handbook- or encyclopedia-writer might with Descartes on mind or soul and its relation to the body.

A first contrast to be noted is that for contemporary dualism establishing the distinction of mental and physical is an end in itself, while for Cartesian dualism it was only part of a larger project. The scope of the project is indicated by the title page of Descartes' most important work, *Meditations on First Philosophy*. Already in Descartes' day a domain of philosophy, which unlike theology may make no appeal to faith, revelation, tradition, authority, or the like, but only to reasoned argument and the evidence of experience, had for centuries been recognized in principle. But in practice arguments that passed for philosophical were often directed towards proving conclusions dictated in advance by theology. A case in point is Descartes' work, as he himself presents it.

For he professed a faith according to which the soul survives death and exists in a disembodied state pending resurrection and re-embodiment. And he maintained that this doctrine is not merely something true and known by faith, but something that, like the existence of God, could and should be proved philosophically as part of a defense of that faith. And the title page to the first, Latin edition of his book promises that in it "God's existence and the soul's immortality are demonstrated."

I am told that this title page comes not from the author himself but from his publisher. Be that as it may, it is false advertising so far as the immortality of the soul is concerned: nothing in the volume even purports to be a proof of that. The title page of the second, French edition is more accurate: "the immortality of the soul" is replaced by "the real distinction

between the soul and the body." But just here is one of the points where Descartes "does not discuss the issues in the same terms in which one would today."

For he does not just mean that soul and body are really distinct, as theology has taught all along. For him "**real distinction**," contrasting with "**rational distinction**," is a technical term, and what is being claimed when a real distinction between two items is claimed is that one could exist without the other. If Descartes, in claiming to prove that the soul is "really" distinct from the body, is not going so far as to claim to prove that it actually does survive the death of the body, he is at least claiming to prove that it possibly could. It is here that the modal notion of possibility comes in.

## Substance or Properties?

Contemporary dualism makes no use of the notion of "substance" in any late medieval or early modern sense. By contrast, Descartes uses, beside the distinction of "real" *versus* "rational," the oppositions of "**essence**" *versus* "**accident**" and of "**substance**" *versus* "**mode**," and other terminology of a kind that sometimes makes the seventeenth century seem closer to the thirteenth than to the twenty-first. For Descartes' argument, in executive summary so to speak, consists of claiming a clear and distinct perception of the truth of three premises which together yield the possibility of mind without body as an immediate consequence: first, that matter is a substance whose essence is extension, not thinking; second, that mind is a substance whose essence is thinking, not extension; and third, that when substances are distinct in essence it is possible for one to exist without the other, which is to say, for God to create one without the other, or having created both, to preserve one without the other.

Since the term "mode" has vanished from the vocabulary of philosophy over the last several centuries, and "substance" as Descartes uses the term only makes sense in contrast with "mode," it is not incorrect to say that few present-day philosophers believe in mental substance in a Cartesian sense. But it is misleading, since few believe in material substance in a Cartesian sense, either. The usage of "substance" in present-day philosophy simply is not Descartes' quasi-scholastic usage. For many it hardly differs from usage outside philosophy.

In colloquial usage, a substance is a kind of stuff as opposed to a kind of thing: the denotation of a mass noun like "gold" or "water," rather than a count noun like "wedding ring" or "ice cube." Contemporary philosophers do not believe in mental substance in such a sense, but then neither did

Descartes. For it is characteristic of substances in this sense to take up space, while this is just what minds do *not* do, according to Descartes.

There *have* been believers in space-filling mental stuff, but Descartes was not of their number. They have included, rather, traditional thinkers who took spirit to be literally a kind of breath, and some in the days of William James who were duped by mediums claiming to produce what they called "ectoplasm," the paradigmatic mental or spiritual substance. They have also included some materialists, notably Pierre Cabanis, if we take literally his assertion that "the brain secretes thought as the liver secretes bile," for secretions in any literal sense certainly do take up space.

Since philosophy has no international body regulating the use of technical terminology, as some sciences do, there is nothing to prevent philosophers today from labeling their views "substance dualism" if they wish. It can only be said that those who do so are using "substance" in a sense different from both Cartesian and colloquial usage. The kind of dualism I want to consider avoids such novel usages.

Contemporary dualism also focuses less on the thinker than on the thinker's thoughts; for Cartesian dualism it is the reverse. This is another part of what may be meant by broad statements about "property dualism" *versus* "substance dualism." The question of the existence of a soul is logically prior to the question of its immortality, and in Descartes' treatment of this prior question, too, the contrast of substance and mode is deployed. His most famous and fundamental insight is that, delusion being a form of thinking, where there is no thinking there can be no delusion, hence it cannot be a delusion that thinking occurs. If it is thought that it does, then *ipso facto* it does. Descartes takes this to imply that "I think" or "*je pense*" or "*ego cogito*" is true every time it is thought.

But where does the I or *moi* or *ego* come from here? It is the background assumption that everything is either a substance or a mode of some substance that leads Descartes to suppose that he can get two for the price of one here, a double conclusion, the existence of both thought and thinker, from a single observation. For if his fundamental insight is that thinking occurs, then since thinking seems clearly not a substance in anyone's sense, his background assumption leads to the conclusion that there must be a substance of which thinking is a mode: a thinking thing or *chose qui pense* or *res cogitans*.

Critics such as the eighteenth century German physicist and wit Georg Lichtenberg suggested Descartes has no right to say "I think" or *"je pense*" or "*ich denke*," but only "it is thinking" or "*il pense*" or "*es denkt*," parallel to "it is raining" or "*il pleut*" or "*es regnet*," where the "it" or "*il*" or "*es*" is a dummy pronoun required by grammar but denoting nothing.

Though it is not a delusion that thinking occurs, the thinking itself may be full of delusions, and one of them may be that in addition to an on-going stream of thinking, there is a continuing, unified subject, the I or *moi* or *ich*, doing the thinking.

That there is a continuing unified subject or "**self**" has been denied by advocates of "bundle" views from Nagasena to Hume and beyond. But this issue does not loom large today in debates between dualism and rival views. Because of the already mentioned shift from claims about mental *components* to claims about mental *features*, contemporary dualism need not spend much time on the issue between Descartes and Hume about the "self." To be sure, thoughts are generally written of as being thought by thinkers, but not much is assumed about the degree of unity and continuity these thinkers enjoy.

If contemporary dualism focuses more on thoughts than thinker, it also focuses more on feelings than thoughts. For Cartesian dualism it was the reverse. Descartes' most general term for mental activity is the one I have been using just now: *thinking*, or its French and Latin equivalents. But Descartes' notion of thinking is in some dimensions broader and in others narrower than ours. He gives a list at the beginning of his third mediation of (much of) what he takes thinking to include. At the top we find affirming and denying, which seem closest to what we today in the first instance count as "thinking."

But as soon as a thinking is mentioned, note has to be taken of a key distinction between **occurrent** and **dispositional**. Suppose I say someone is thinking that COVID has become endemic on the grounds that just now he is saying, if not aloud, then at least to himself, "COVID has become endemic." Then I am speaking of occurrent thinking. Suppose instead I say someone thinks that COVID has become endemic on the grounds that this is something she tends to say if the topic comes up, though just now she is not saying anything, even to herself, but sleeping dreamlessly. Then I am speaking of dispositional thinking. English happens to have a difference in verb form, "is thinking" *versus* "thinks," to mark this distinction, but French, for instance, does not. Nonetheless we can be confident that when Descartes writes of affirming or denying, it is occurrent thinking that he is writing of.

Present-day dualists' most general term for the kind of mental states they wish to consider is **conscious**. But it is a commonplace that, most of our language being designed for speaking about things in our shared and objective external world, attempts to discuss our personal and subjective internal worlds may very often end in miscommunication. The term "consciousness" is a case in point, being involved in many merely verbal

(though also in many quite substantive) disagreements.

An amusing instance is provided by the work of Julian Jaynes (1976) with his doctrine of what he calls the "bicameral mind." According to Jaynes's mind-boggling hypothesis, some three millennia or so ago people did not deliberate but rather, when they faced a crucial decision, would hear voices, coming from the right side of the brain but interpreted as divinities, telling them what to do; and what he calls "consciousness" arose only when the voices fell silent. On the usage of "conscious" adopted here, this is a nonsensical claim. If people were hearing voices, they were having auditory hallucinations, which are a kind of auditory sensation, which are a kind of conscious experience; and so they were already conscious. Jaynes has fallen into writing "conscious" where on the usage adopted here one would have to say something like "**self-conscious**."

Because of variations in the use of "conscious," one cannot *just* say, "Let us restrict our attention to conscious states," and expect to be understood at once. Rather, one had best offer both an alternative characterization of the general class of states in question, and a representative list of instances. The most commonly cited characterization of conscious states is that of Thomas Nagel (1974): They are states that "there is something it is like" to be in. Stevan Harnad would amend this to "something it feels like."

Dispositional thinking will be set aside henceforth, since it is not conscious in this sense. In dreamless sleep there is no feeling, and hence there is nothing it feels like to be dreamlessly sleeping, nor anything it feels like to be in a state that one can properly be said to be in even when dreamlessly sleeping; and this includes dispositional thinking. But in our day, in contrast to Descartes' time, it is widely recognized that there exist many kinds of *occurrent* unconscious thinking. A large part of cognitive psychology and philosophy of mind are concerned with activities or processes going on below the level of consciousness, but nonetheless perfectly appropriately called "mental."

Materialism makes a universal claim, that *everything* is physical. By contrast, dualism makes only an existential claim: *some* mental states are not physical states; more specifically, conscious states are not. But in our day, unlike Descartes' time, the dualist has to recognize that the conscious states may be only a small fraction of the mental states of interest in the science and philosophy of mind. That is why I said at the outset that work on the mind-body problem "constitutes only a restricted and rather specialized part of a large and growing field." Still, in what follows, unless explicitly indicated otherwise, "mental" will be tacitly understood as short for "*conscious* mental," that being the side of mentality involved in the

mind-body problem that is our topic.

When it comes to a representative list of instances, Descartes begins, as I said, with affirmation and denial. At the bottom of his list comes what he calls "perception" in a broad sense. It is presumably here that one should place the states contemporary dualism places at the top of *its* list. These including both **sensations** (such as warmth) and **affects** (such as euphoria). In colloquial English such are often called "**feelings**," an expression used for all affects and all senses except the four that have their own terminology: taste, smell, hearing, sight. But in fact much of the literature is concerned with sight or visual sensation, and more specifically with color sensation, and more specifically still with the experience of seeing red.

Conscious, occurrent thinking, thinking there is something it is like to be engaged in, *is* also on the present-day dualist's list, but perhaps near the bottom. Low priority may be given even to consideration of the many perceptual or emotional states with a lot of thought content to them, such as seeing the traffic signal change to stop or worrying that one will be late to the meeting, in contrast to seeing something red and shining or feeling free-floating anxiety, which are closer to being "raw feels."

For the most part **intentionality**, the feature of being directed towards an object, a feature perhaps most clearly exhibited in "thoughts" of the kind we call beliefs and desires, will not be much discussed, since it is not specific to conscious mental states, being shared with many unconscious ones. Moreover it seems a feature only of *many* conscious states, but not *all*. For instance, it is not easy to locate an intentional element in affects such as euphoria. There is an *analogue* of the mind-body problem about where intentionality fits, if it does fit, into the physical cosmos; this will not be our problem here.

Contemporary dualism has been much concerned with the possible existence of bodies—living ones, not corpses or cadavers—without minds, where Cartesian dualism was more concerned with minds without bodies: souls of the departed, or more crudely, **ghosts** in a philosophical as opposed to a folkloristic or horror fiction sense. If what is common to the two dualisms were summed up in a single formula, it might be "mental and physical are distinct, *with no necessary connection between them*." Modal logic teaches that the absence of a necessary connection is equivalent to the presence of a possibility of disconnection. Disconnection could involve either bodiless minds or mindless bodies (or both). The first possibility is the Cartesian option, the second possibility is the contemporary option, though how it came to be so is a longish story that remains to be told.

# CHAPTER 2

# NEURALISM

## Minor Forms of Materialism

Fifty years or so ago Saul Kripke shook up the anglophone philosophical world with a series of three lectures at Princeton, subsequently published (1972) as *Naming and Necessity*. That work has become required reading for students of philosophy, and I suppose most readers are familiar with it. I have given my own account elsewhere (Burgess, 2013), and have no need to summarize it at this point, though something will have to be said later. The lectures culminated in a discussion of the mind-body problem. Kripke did not argue for a particular position (saying instead that the problem is "wide open" and "very confusing"), but was critical of materialism, or at least the way its adherents were formulating it at the time. Let us look first at that formulation, before taking up Kripke's intervention.

Materialism has been defined so far as the doctrine that everything is physical. That won't quite do if one recognizes a category of **abstract** entities, a "third world" as far from the mental as it is from the physical, containing among others the objects of mathematics. For to take everything to be material would be to deny the existence of such objects, thus embracing what is called **nominalism**, and embroiling the materialist in debates over philosophy of mathematics far indeed from the mind-body problem. So let us say instead that materialism takes everything *mental* to be physical. Since we are dealing now not with "substances" but with "properties" (or events or facts or states or whatever), the thesis must be that every mental property (or whatever) is a physical property (or whatever). Since properties and the rest are themselves often cited as paradigmatic abstract entities, what is understood by a "physical property" cannot be a property that is a physical entity, but rather must be a property pertaining the physical side of things—a formulation obviously in need of further elaboration, which it will be given later.

Dualists might cite sensations or affects as counterexamples to materialism. One extreme form of materialism, called "**eliminativism**,"

would simply deny that these exist or occur, claiming that it is "vacuously" true that everything mental is physical simply because there *is* nothing mental: there *are* no feelings, and further, contra Descartes, to think that thinking is occurring would be a delusion. A reader new to the subject might doubt if anyone could really hold a view so silly, but there have historically been many examples of philosophers talking themselves into views that are even sillier.

And some contemporary materialists do at least *write as if* they embraced eliminativism. For instance, it has often been suggested that Daniel Dennett's 1991 book *Consciousness Explained* ought to have been entitled *Consciousness Denied*. But when critics point to passages in certain materialists' works where they seem to deny the existence consciousness or sentience, these materialists generally can be found pointing to other passages where they after all acknowledge that thought and sensation do occur. But then critics may accuse them in these other passages of engaging in "**Humpty-Dumptyism**," redefining ordinary words to mean something else. The interested reader can find a classic example of this dialectic in an exchange a few years back between Dennett and Galen Strawson (beginning with Strawson 2018, and followed up in later issues of the same periodical; see also the earlier Searle 1995 and Nagel 2017).

In an earlier draft of this work I gave the eliminativist position some fairly extended discussion, but I was advised by experts that this was unnecessary, since "no one" (which I take to mean, "no great number of prominent philosophers specializing in the area") holds that sort of view any more. I had feared otherwise, especially after hearing a noted psychologist at my own school compare the belief that we are conscious to the delusions of a psychiatric patient who maintained there was a squirrel inside his skull. I am relieved to be told authoritatively that eliminativism may be ignored, and complying with this advice I will henceforth gladly ignore it.

The more common form of materialism may be called **assimilationism**. It allows that there are items that can properly be called "mental," but insists that these can all with equal propriety be called "physical," and more specifically "neural" or nervous-system-related, and usually more specifically still "encephalic" or brain-related. But even this view comes in two forms. One much-discussed around the time of Kripke's lectures, when it was a novelty, is the "**anomalous monism**" of Donald Davison (1970), according to which each **token** mental event, say Don's becoming angry right now, is identical to a token physical event, perhaps Don's developing an excess of bile right now, but no **type** of mental event, such

as becoming angry, is identical to any type of physical event, such as developing an excess of bile.

Now hardly any suggestion in this area could be less plausible than to hold that mental events are identical with physical events one-by-one, but in a lawless, haphazard way, as if Don's becoming angry right now were his developing an excess of bile right now, but Donna's becoming angry yesterday was her developing a deficiency of phlegm back then. This is not really Davidson's view, it turns out. He does allow there can be type-type laws after all, but insists that these can never have the degree of exactness that many philosophers would attribute to the laws of fundamental physics.

To this claim it has frequently been replied that the laws of other sciences—physiology and medicine would be the most relevant to philosophy of mind—do not have the exactness of laws of fundamental physics, either. (For that matter, neither does much of actually existing physics.) Even in chemistry, now very closely indeed integrated with physics through the quantum theory of the chemical bond—so closely that in simple cases the energies of chemical reactions can be predicted *ab initio* by calculation from fundamental physical principles—there is no reduction in the sense of exact definitions of such chemical concepts as "covalent bond" or "ionic bond" in fundamental-physical terms. What the quantum theory shows, rather, is that these chemical notions are to a degree imprecise.

In any case, the "mental" items with which Davidson is concerned are not "raw feels" but rather such **propositional attitudes** as instances of belief or desire or the like, which we have already set aside. Moreover his interest is not in the first-person attribution of such attitudes to ourselves when we are conscious or aware of them, but their third-person attribution to others, which on his view always involves "radical interpretation," to make those others comprehensible in our own terms. Thus Davidsonian anomalous monism is not a view on the mind-body problem at all, when that problem is understood as it is here.

## Two Dualist Concessions: Correlation and Uniformity

The dominant version of assimilationism circa 1970, and Kripke's apparent target, was rather what followers of David Armstrong (1968) called the "central-state identity theory," or simply **identity theory** for short, identifying each *type* of mental property (or event) with some *type* of neural property (or event). To avoid confusion with other senses of "identity," I will call this view **neuralism**. Not only was it widely held

among materialists, but even dualists were generally prepared to concede, at least for the sake of argument, the **psychophysical correlation principle**, according to which types of mental items are at least *correlated by laws of nature* with types of neural items. The disagreement was that materialists took the correlation in question to be of the closest or strongest kind, outright identity, while dualists took it to be something less than that.

It cannot be claimed that the correlation principle is an established scientific result comparable to, say, the conservation of electrical charge. And what makes the principle plausible is by no means our having discovered the exact laws of correlation operative in many cases, or even any. Our present-day lay and scientific language may not yet even contain all the mind-related and brain-related terminology ultimately required for the formulation of any pertinent laws; indeed, it almost certainly doesn't. Such plausibility as the principle has is the result of our having in various cases gotten what seem closer and closer approximations to an ultimate law as we have learned more and more about the neuronal mechanisms connected with one or another phase of conscious experience.

The fact that no generally accepted exact law is known makes it impossible to give an uncontroversially real example of a correlation. For the sake of giving examples it is customary to pretend that we *do* after all know the law in one important case. The pretense is that we know the neuronal correlate of the state of being in **pain** to be the state of having **C-fibers** (or nerve-fibers of the C-bundle) that are firing. One says "pain = C-fiber firing" for short.

This is *only* a pretense because, for one thing, other kinds of fibers are involved with pain (and sharp so-called first pains, as contrasted with lasting second pains or aches, are said to involve the firing of C-fibers less than that of Aδ-fibers), and for another thing, C-fibers are involved in sensations other than pain (and the firing of histamine-selective C-fibers is said to be involved less with pain than with itching). But because the example is only a pretense, no actual knowledge of what C-fibers are or do is required to follow its typical role in debate. Another equally over-simplified, pretend or toy, example would take the correlate of euphoria to be an abundance of endorphins in the brain.

So much for the moment about the correlation principle. It is a dualist concession to materialism that may be regarded as an elaboration of Descartes' view that in living human beings, the brain is the part of the body most closely linked to the mind. By contrast, a further concession also made by many dualists directly conflicts with Cartesian physics, and this I must now attempt to describe.

Descartes, though much greater as a mathematician than as a physicist—Cartesian coordinates are today taught to middle-school students around the world, while Cartesian vortices are remembered only by specialists in the history of seventeenth-century protoscience—did have a role in laying out the program for the new physical science of his day. He was among the founders or pioneers of the project of attempting to explain many physical phenomena in terms of a few mathematically-formulated laws; and he more or less foresaw that there should be among these some kind of law of inertia and some kind of law of conservation. He never got the right laws, mainly because he lacked the Newtonian concept of mass. There is a certain picture, or caricature, of this aspect of Descartes' view that I am told derives largely from G. W. Leibniz and his criticism of Cartesianism among his contemporaries, a picture according to which even if he *had* had the concept of mass, Descartes could not have accepted basic principles of Newtonian mechanics such as the **conservation of momentum**.

On this picture, the law of the conservation of momentum, the product of mass times directed velocity, cannot be admitted by a Cartesian, but at most a law of the conservation of the product of mass times undirected speed. A Cartesian might grant that conservation of momentum may hold throughout the mineral and vegetable kingdoms, and among the so-called lower or non-human animals, and in such parts of the human body as the heart and liver and spleen and kidneys, and even in most of the brain—but not in the pineal gland. At least the direction if not the speed of the motion of the fluid supposed to fill this part of the brain can fail to be conserved, because the flow or motion of the material fluid is deflected and redirected by action of the immaterial soul. On such a view, laws of nature governing physical systems may have exceptions in the special case where the system in question includes this one crucial part of the human brain.

Today, not only is the correlation principle widely assumed, but a further **uniformity principle** seems to be almost equally widely, albeit perhaps less explicitly, assumed, according to which matter and energy are everywhere subject to the same physical laws, even in systems of which a human body or part of one is a component. And where the laws are statistical rather than deterministic, they apply inside and outside the body and brain with no change in the statistics. All present-day research in neurophysiology, in particular, seems to take something like this for granted. No one seems to be seriously looking for violations of or exceptions to the laws of physics in the pineal gland or elsewhere inside the skull, not even those such as Francis Crick (1995) who sometimes write as if they hope to find the soul there.

The content of the uniformity principle may be made clearer by an analogy. Consider spherical bodies of a fixed size approximating that of a ping-pong or golf ball, constrained to move on a frictionless horizontal track one end of which is arbitrarily designated "left," the other "right." Velocity in the case of such constrained motion may be understood to be simply speed with a plus sign in the case of motion from left to right, and a minus sign in the case of motion from right to left. Now if spherical bodies $x$ on the left and $y$ on the right are moving with velocities $a$ and $b$ respectively, and if $b < a$, the two will collide, and thereafter move apart with some velocities $c$ and $d$ for $x$ and $y$ respectively. This description is purely **kinematic**, involving only size, shape, speed, and direction, but no **dynamic** quantities such as mass or others defined in terms thereof, such as momentum or energy.

Allowing for idealization, *some* of the laws that govern such cases can themselves be stated in purely kinematic terms. For instance, if in-coming velocities $a$ and $b$ for a certain pair of spherical bodies are followed by out-going velocities $c$ and $d$, then in a repeat of the experiment, the same spherical bodies with in-coming velocities $a + t$ and $b + t$ will have out-going velocities of $c + t$ and $d + t$ (as per what is called "**Galilean invariance**"). But there are other laws that govern such cases that can only be stated by bringing in dynamic quantities. For the $c$ and $d$ following upon the same $a$ and $b$ will not be the same in the case of two ping-pong balls as in the case of one ping-pong ball and one golf ball. To predict $c$ and $d$ one must take into account also the masses of the bodies involved.

In this sense, a class of situations described in purely kinematic terms may be subject to laws of nature with only some and not all of the pertinent laws being stateable in purely kinematic terms. The uniformity principle in effect says that the relation between the physical and the mental is *not* like the relation between the kinematic and the dynamic in this respect. As has been mentioned, sensed color was excluded in the seventeenth century from the list of features of objects on which physical science was henceforth to focus. What the uniformity principle in effect says is that such features will never have to be brought back in, at least not in order to state the underlying laws governing aspects of situations describable in purely physical terms, in the way mass has to be brought in to explain the kinematically-describable behavior of colliding balls.

For another analogy, there was a time when many workers in botany and zoology assumed that matter may be affected in the case of all kinds of living things from microbes on up by an immaterial "life force" or "*élan vital*." This view, **vitalism**, has been thoroughly discredited for over a century. The uniformity principle goes a step further than the rejection of

vitalism by disallowing exceptions to the laws of physics even in the special case of conscious human beings.

The uniformity principle may be described as denying **hard interactionism** or **interventionism**, the view that mind or spirit can so interfere with or intervene in the material world or physical cosmos as to divert or deflect a course of events heading in one direction so that it follows a different direction instead (as his interaction with the ghosts diverted the course of Scrooge's life, which was heading towards a bad end, towards a better one). Something like hard interactionism was upheld not only by Descartes but by many down to the time of William James, who showed sympathy both for **hard libertarianism**, which claims for human free will just such a power of diverting or deflecting the course of a life or of history, and also for "psychical research," now known as **parapsychology**, which aimed to come up with reliably reproducible results demonstrating the interference or intervention of the mental or spiritual in some way transcending ordinary limitations on the physical and material.

Today, parapsychology seems nearly as thoroughly discredited—in scientific circles: I am not speaking of popular culture—as is vitalist parabiology. Sporadic miracles cannot be proved not to occur (though the skeptic will insist that the burden of proof lies with those who hold that they do, not with those who hold that they don't); but miracles aside, it is widely held among philosophers today that there is no interference with the physical by the aphysical that is regular and systematic enough that it would have to be taken account of in a scientific psychology. No credible case has been made for, say, the ability to transmit something from one brain to another without any physical signaling (of familiar kinds or of the kind involved in stage-magicians' tricks), or the ability by thinking about them to alter the probabilities of random quantum events. This last was the object of extended efforts at my own university's now-defunct Engineering Anomalies Research Laboratory. I share the sense of some of my colleagues that the extended operation of this laboratory was an embarrassment to our school, and I will accordingly here take the uniformity principle for granted.

I have formulated the principle in terms of natural law, or in fancier jargon, in a **nomological** version. When the topic is discussed in the literature, especially in the wake of the work of Jaegwon Kim (2005), the issues instead are usually formulated in terms of the notion of causation. Then the widely accepted doctrine, corresponding to the uniformity principle in the nomological formulation, is that of the **causal closure of the physical**: any physical event that has a cause has a physical cause, a

principle understood to apply to statistical causes as much as to deterministic causation. But now a division appears.

Some dualist adherents of the uniformity principle, or **uniformitarians** as we may call them, draw the conclusion that mental events are not causes of physical effects in any interesting sense of "cause," a view known as **epiphenomenalism**. An early advocate was "Darwin's bulldog" T. H. Huxley, who compared the relation between body and mind to that between a steam locomotive and its steam whistle, which has no effect on the motion of the train.

Other uniformitarians will endorse **weak interactionism**, upholding causal efficacy of the mental in *some* important sense of "cause," though not one implying that minds or spirits can overcome or override physical law. In doing so they implicitly embrace **overdetermination**, whether they use that term or not, in effect assuming that an event that already has a fully sufficient physical cause may also have a wholly redundant concurring mental cause (or in fancier versions, so redefining "causation" that having certain kinds of associated physical causes *counts as* or *constitutes* having mental causes).

The situation is similar to that in the free will debate, where those who reject hard libertarianism divide into **hard antilibertarians** or **determinists**, who hold that we do not have free will in any interesting sense, and **soft libertarians** or **compatibilists**, who hold that there is some weaker-than-traditional but still interesting sense of "free will" in which we can be said to have it. The former correspond to the epiphenomenalists, the latter to the overdeterminationists.

Obviously there is a danger of the dispute between epiphenomenalism and overdeterminationism degenerating into a purely verbal disagreement and haggling over the meaning of the word "cause," the phrase "make a difference," or some other, related vocabulary. That is one reason I have preferred to stick with a formulation in terms of natural law, and a label "uniformitarian" that will cover both 'isms—though of course "law" itself has been variously understood almost as much as "cause." Another reason is an impression that the divide between Cartesian or Jamesian belief in interference or intervention and either of the two uniformitarian views is more significant, historically and intellectually and culturally, than any disagreement between the latter pair.

I more than once heard my late colleague David Lewis say that to the Victorians, *all* present-day positions in philosophy of mind would have been considered forms of materialism. What I *think* he meant that to those who truly believe, as many Victorians and Edwardians did, that the crucial issue is that of the power of mind or spirit to deflect or divert the

development over time of the material or physical world, changing the course of history, all views that deny such a power are much of a muchness.

As Marx's bulldog Friedrich Engels, an avowed materialist, in *Socialism: Utopian and Scientific* described Huxley's view as "shamefaced materialism," so an avowed epiphenomenalist might describe overdeterminationism as "shamefaced epiphenomenalism." And shamefaced or shameless, epiphenomenalism does not differ from outright materialism as we have been understanding it in a way that would lead to different answers to the question whether spirit can redirect the course of events in the natural world, whether by redirecting the motion of fluid in the pineal gland or in some other way.

Yet here in treating the differences between epiphenomenalism and overdeterminationism as of secondary importance my reading suggests I may be expressing a minority view. For I have found that in the literature, showing that a theory leads to epiphenomenalism seems by many to be considered outright checkmate, while showing that a theory leads to overdeterminationism is considered at most giving check. But let me relegate any further discussion of such matters to an optional Appendix.

## Preliminary Skirmishes

It is common in debates over the mind-body problem for arguments that seem definitive and dispositive to one side to seem feeble and flimsy to the other. This is so to a degree that raises suspicions in some, notably William Lycan (2009), that allegiance to dualism or materialism may be more a matter of personal temperament than of anything else. Perhaps it is in Jamesian terms a matter "**tough-mindedness**" and "**tender-mindedness**." Certainly such contrasting temperaments can be found at work in reactions to the simplest arguments pro and con neuralism, where disagreements can pretty quickly come down to what are apparently irreducible and irreconcilable "intuitions."

The simplest case *for* neuralism rests simply on **Ockham's razor**, the precept that entities are not to be multiplied beyond necessity. Science is obliged to recognize neural states or events, but if, as dualists allow, mental states or events correspond perfectly to these, there is no reason to posit, in addition to the physical, such strange further aphysical entities. To this the dualist reply is that while perhaps entities should not be recognized beyond those it is *necessary* to recognize, still it *is* necessary to recognize what is staring one in the face, and nothing stares us more directly in the face than sensations and affects. Moreover, nothing is more

familiar; which is to say, nothing is less strange.

The simplest case *against* neuralism points out that the discovery of psychophysical correlation, let alone alleged identity, is thoroughly ***a posteriori***, dependent on experience, rather than ***a priori***, independent of experience. It is only through experience that we learn that sensations and affects have physical counterparts or correlates, and that these are located not in the clouds but in the body, and within the body not in the heart or liver or spleen or kidneys, but in the brain, and further in certain localized regions or specialized activities thereof.

In order to establish the correlation between, say, pain and C-fiber firing, we must have had some way of recognizing when pain is present that is not dependent on the antecedent assumption of this correlation; and what way of recognizing the presence of pain could this be, other than by *feeling* it, or accepting others' reports of doing so? But experience of this feeling, on which any claim of correlation, let alone identity, must rest, is left out of the neuralist's ultimate account. To this the neuralist reply is that the feeling of pain is *not* left out, since their account includes the firing of C-fibers, and that, they claim, is quite literally one and the very same thing. (Once the dualist has conceded the correlation principle, any claim that something has been left out can always be answered this way—I will not say how plausibly.)

And now we come to the peculiarity of the way in which this claim of the *a posteriori* character of the discovery of the relation between pain and C-fiber firing would have been expressed circa 1970. At that date, it was at least sometimes recognized that there is at least a *conceptual* distinction, or difference in "**intension**," between two contrasts: the epistemological contrast of the *a priori* with the *a posteriori*, what can be known independently of experience and what cannot, and the "metaphysical" contrast of the **necessary** with the **contingent**, of what is and could not have been otherwise with what is but could have been otherwise. It was nonetheless widely assumed that the two distinctions agreed in "**extension**," with everything *a priori* being necessary and vice versa.

Kripke began his lectures by illustrating how many philosophical writers had identified—or as he would claim, confused and conflated—the two distinctions. In particular, the alleged coincidence of pain and C-fiber firing was often described, on account of its being an *a posteriori* discovery, as a discovery of something "contingent": the coincidence of pain and C-fiber firing was claimed to be a "contingent identity." This is the formulation that Kripke most centrally opposed, and that in the wake of his work has largely disappeared from the literature. (No philosophical view, having been stated, ever goes away completely.)

He gave many examples, which a large part of his audience and readership eventually found convincing, of alleged *a posteriori* necessities, and he argued in particular that "**scientific identifications**," as that of water with $H_2O$, or heat with random molecular motion, were of this character. Similarly, if the identification of pain with C-fiber firing were true, it would have to be necessarily and not just contingently true, another example of an *a posteriori* but non-contingent identity.

This observation, if accepted, hints at a new route by which a dualist might attempt to establish the distinction between pain and C-fiber firing: To show that the two are *actually* distinct, it would be enough to show that they are *possibly* distinct. If they *could have been* distinct, then they actually *are* distinct; and similarly for other psychophysical correlations. For instance, if the neuronal correlate of seeing red *could have been* present without the sensation of seeing red, then the two *are* distinct. If the existence of a being physically like us who saw an **inverted spectrum** (like the film negatives of old-fashioned color photographs) is possible, then while it may be that no such being actually exists, still the distinction between the color sensation and its neuronal correlate would follow.

Since David Chalmers' *The Conscious Mind* (1996), a landmark that came out about midway between Kripke's lectures and the time of this writing, the case that has been most discussed is that of beings physically like us—moving around and emitting sounds just as we do when pursuing our desires or expressing our beliefs—but not just with a mental life different from ours, as in the inverted spectrum case, but with *no (conscious) mental life at all*. Such are **zombies**, in a philosophical as opposed to a folkloristic or horror fiction sense. The boldest dualist hypothesis has become this, that there could have existed a whole **zombie world**, physically like our world but without any (conscious) mental life at all going on in it; a world governed by the same physical laws as ours, but without any psychophysical correlation laws at all. The Kripkean argument would then show that if a zombie world is so much as possible, then no (conscious) mental state or event is identical with any physical state or event.

But what Kripke gives with one hand, he seems to take away with the other. Since he has distinguished that which we may call the **conceivable** in a technical sense of not being *a priori* false, from that which we are calling the **possible** in the sense of not being necessarily false, to show the possibility of a zombie world it will *not* be enough just to show the conceivability of a zombie world. That there may be a new, modal route to dualism has been suggested by the principle that possible difference implies actual difference; that the route from the acknowledged conceivability