

## Russellian identity theory and neutral monism

Thus far in this book we have focused on dualism and materialism as the main alternative general metaphysical approaches in the philosophy of mind. That is, we have considered the views that everything is ultimately material (materialism), and that the material and the mental are equally ultimate (dualism). These alternatives are paid the most attention by contemporary philosophers of mind, but they are not the only alternatives to be proposed in the history of the subject. A third view, known as idealism, holds that everything is ultimately mental - for example, the version associated with George Berkeley (1685-1753) holds that purportedly physical objects like tables and chairs really exist only in so far as a mind perceives them to exist. But though idealism has had some illustrious defenders in the history of philosophy, it is not generally regarded as a serious option by most contemporary philosophers (with some important exceptions). There are two other, more promising, alternatives that we will be exploring, one in this chapter and the other in chapter 8. The first holds that neither mind nor matter is metaphysically ultimate: what is ultimate is rather a single kind of stuff that is neutral between, and more fundamental than, either of them. This is, in a nutshell, the metaphysical theory known as neutral monism.

The most important proponent of this view in the twentieth century was Bertrand Russell. His formulation of it evolved significantly through the course of his long career; what we want to focus on is the final, settled version. Russell begins by drawing out the implications of the indirect realism he endorsed, and which we discussed in chapter 1. If in perception we are directly aware, not of external physical objects themselves, but rather only representations of those objects, then we have in Russell's view no grounds for supposing that those objects really have the properties they are presented to us by perception as having. We have no reason to assume, for example, that

*Consciousness 99*

the redness and sweetness of the apples we perceive is really in the apples themselves, as opposed to being merely an artefact of our perceptual machinery - just as the redness you see on the wall in front of you when you are wearing glasses with red lenses is, for all you know, not really in the wall itself but only an artefact of the glasses. As we've noted before, physics seems to give us positive reason to believe that the redness and sweetness are not in the apples: for like every other physical object, an apple is in reality nothing but a collection of colorless, odorless, tasteless particles. What the physical world is really like "in itself," apart from our perceptual representations of it, is not something perception can tell us.

What does tell us what the physical world is really like is science. But science, Russell argues, does not tell us nearly as much as we often assume it does. For instance, what exactly are these colorless, odorless, tasteless particles of which physics speaks - molecules, atoms, quarks, gluons and so forth? Physics defines these entities entirely in terms of their causal relations to one another: a molecule is whatever plays such-and-such a causal role at the microscopic level, an atom is, among other things, what plays the role of serving as a component of a molecule, and so on. But what exactly it is that happens to play these roles is something physics does not tell us. We know from science only that the material world is a collection of fundamental entities having a certain causal structure, a structure described in mathematically precise detail by the physical sciences; but what it is that fleshes out this causal structure, the intrinsic nature of the specific entities that bear these causal relations to one another by filling out each place in the vast causal network described by science, is something we do not know. (This is a view about the nature of scientific knowledge known as structural realism: realist because it holds that there really is a physical world existing external to our minds, structuralist because it holds that all we know of that world is its structure rather than intrinsic nature.)

Our knowledge of the external physical world turns out to be highly abstract; including our knowledge of the brain, considered as the object of neuroscientific research, as one external physical thing among others. The brain is not in reality the greyish, squishy thing we encounter in perception:

that is only a subjective, perceptual representation of the brain. The brain is, rather, a complex causal structure of neural events, where these neural events are defined in terms of their characteristic causes and effects rather than in terms of the qualities presented to us in visual or tactile inspection of the brain. The inner nature of what specifically has these cause and effect relations is something we do not know –

### 100 *Philosophy of Mind*

or at least, we do not know from either perception or neuroscientific study.

But are perception and scientific inquiry (whether neuroscience, physics, chemistry or whatever) the only possible sources of knowledge about the nature of the brain? Russell suggests that there is one further possibility: introspection. In introspecting or looking within itself, the mind is directly aware of its own contents - of thoughts, experiences, and their associated qualia. As materialists have argued, there are, at least in general, correlations between various mental events on the one hand and brain events on the other. Perhaps in introspecting these mental events, and in particular our qualia, we are directly aware of precisely the inner natures of the entities that play the causal roles specified by neuroscience. Perhaps neural events just are the thoughts, qualia, and so forth encountered in introspection. In being immediately aware of the taste of an apple or a sensation of pain, maybe what we're directly aware of are events occurring in the brain, as it really is "in itself."

This is obviously a mind-brain identity theory. But it is not the materialist kind of identity theory discussed in chapter 3. Materialism in general seems to take it for granted that we know exactly what the intrinsic nature of the physical world is, and seems to assume also - especially in the case of functionalism - that we do not know (or at least that pre-philosophical and pre-scientific common sense does not know) what is the intrinsic nature of the mental realm: the functionalist claims that mental states and processes are to be defined entirely in terms of their causes and effects. Russell's view is that this has things precisely backwards. It is in fact the mental world that we know most directly and intimately, and the external physical world that we grasp only in terms of its causal structure. In identifying the mind and the brain, Russell is not, as the materialist identity theorist is, reducing the mind to the brain; if anything it is the other way around. The brain turns out to be the mind; more exactly, the neural events and processes defined only abstractly, in causal terms, by neuroscience turn out to be nothing other than mental events and processes - thoughts, experiences, and the like. The grey squishy thing you've seen pictures of in textbooks or that a neurologist looks at when doing surgery is not what the brain is really like intrinsically. If you want to know what it is really like, you need only focus on the qualia you're experiencing right now. The whiteness and blackness of the paper and ink of the book you're reading, the colors on the cover, the smell and warmth of the coffee in the cup beside you, the feel of your back against the chair: those are the

*Consciousness 101*

brain's true qualities. In introspecting those qualia, you are directly aware of nothing other than the inner nature of your own brain. or, as Russell paradoxically put it: "I should say that what the physiologist sees when he looks at a brain is part of his own brain, not part of the brain he is examining"!

If this sounds strange, it is supposed to. But it makes perfect sense when one combines indirect realism with the mind-brain identity thesis. For what Russell means is that the physiologist is not directly aware of the (patient's) brain he is examining, though of course he is aware of it indirectly; what he is directly aware of is a constellation of qualia - greyishness, squishiness, etc. - which are, given the identity theory, identical to features of his own brain, and which are ultimately a distant effect of the light reflected from the patient's brain traveling to the physiologist's retinas, which sets up a sequence of neural firing patterns eventually culminating in the visual experience. still, the theory definitely counts as a revision of common sense. More importantly, for our purposes, it counts as a rejection of materialism, for, both epistemologically and metaphysically, it gives priority to the

subjective, first-person realm of qualia rather than the objective thirdperson external physical world. Yet it also seems to count as a rejection of dualism, in so far as it identifies the brain with the mind, rather than seeing them as distinct substances.

Indeed, it might seem at first glance to lead instead to a kind of idealism: for if qualia are the intrinsic qualities of the brain, and the brain is - as far as we know from science - made of exactly the same kind of stuff as everything else in the physical universe, wouldn't this entail that everything else in that universe also has qualia as intrinsic qualities? wouldn't qualia be what ultimately make up tables, chairs, rocks, trees, and every other object of everyday experience? If so, this would seem to entail that, in some sense, everything physical is really mental, which is precisely what idealism claims. But Russell and some other philosophers who have endorsed and developed his position, such as Michael Lockwood, have resisted this conclusion. They have suggested that what contemporary philosophers have come to call qualia (this was not Russell's own expression) - reddishness, the nagging character of pain, the pungency of an odor - may well indeed be the intrinsic properties of every physical thing; but they have also suggested that these properties are, contrary to the standard view, not in fact essentially mental properties at all. Reddishness and all the rest need not necessarily exist in the mind of an experiencing subject: they can exist unsensed by any mind, and do so exist when they enter into

## 102 *Philosophy of Mind*

the constitution of physical objects other than the brain. The Russellian view is thus interpreted - at least by Russell himself and Russellians like Lockwood - as a version of neutral monism: qualia comprise the single ultimate kind of stuff out of which everything in the world is composed (hence "monism"), but they are intrinsically neither mental nor non-mental (hence "neutral"); they count as mental only when organized into the sort of causal structure described by neuroscience (that is, a brain), and count as non-mental when Organized into other sorts of causal structures (rocks, trees, tables, chairs, galaxies). Since it identifies qualia with properties of the brain, this account is also a kind of identity theory - sometimes labeled the Russellian identity theory, to distinguish it from materialist identity theories of the sort described in chapter 3.

One of the advantages of this theory, whatever one wishes to call it, is that it seems to be immune to the sorts of objections that, as we've seen, plague materialist theories. In response to the zombie argument, for instance, the Russellian can hold that zombies can be shown not truly to be conceivable when one's exercise in conception is informed by indirect realism (and the structural realism Russell conjoins to indirect realism). Zombies seem conceivable only if, when imagining them to be "physically identical to us," we imagine their brains being the greyish, squishy things we encounter in perception. But of course, to imagine that sort of thing is really only to imagine a perceptual representation of a brain; it no more involves imagining the brain as it really is intrinsically than does imagining a linguistic representation like the word "brain." To note that a greyish, squishy thing can be imagined to exist apart from qualia no more undermines a mind-brain identity theory than the fact that you can imagine the symbol "H<sub>2</sub>O" existing in the absence of water undermines the claim that water = H<sub>2</sub>O. Really to imagine the brain as it is "in itself" would, on the Russellian view, require imagining it as constituted by qualia. But to imagine that is, by definition, not to imagine a zombie, since a zombie is supposed to be a creature devoid of qualia. In that case, however, zombies turn out to be inconceivable after all.

### **Troubles with Russellianism**

Or do they? A number of philosophers take the Russellian position - long neglected in the philosophy of mind, but in recent years making something of a comeback - to be a great advance over the standard alternatives. But arguably, it will not do as it stands. First, the

suggestion that qualia can exist independently of any experiencing conscious subject is highly counter-intuitive, indeed highly implausible. The very notion of qualia is, after all, introduced as the notion of properties of immediate conscious experience. So it is questionable whether we can coherently abstract away from the notion of qualia the presence of a conscious subject, a mind, to whom they are presented.

Some philosophers sympathetic with the Russellian approach, such as David Chalmers, acknowledge that qualia require a conscious subject for their existence - and thereby accept the idealism (or panpsychism, as they often prefer to call it, to distinguish their view from the sort of idealism associated with Berkeley) to which this commits them. They don't hold that qualia quite like ours - pains, itches, color sensations, odors, and the like - make up the physical universe outside our minds, for our qualia are no doubt more complex, given the complexity of our brains. At the level of molecules, atoms, and subatomic particles, there are instead what might be called proto-qualia playing the relevant causal roles, properties simpler than, and only vaguely analogous to, our qualia. Associated with these proto qualia, and thus with molecules, atoms, and subatomic particles, would have to be proto-subjects - simple, tiny minds (or proto-minds) having extremely simple experiences (or proto-experiences). It is only when these proto-qualia get organized into highly complex structures like our nervous systems that they somehow, in combination, give rise to complex minds like our own.

The initial, uncharitable objection to all of this is that it is just plain crazy, and Chalmers' critics have not been shy about raising it. For most philosophers, if a theory has implications as bizarre as that basic physical particles are associated with minds (proto- or otherwise) experiencing qualia (proto- or otherwise), that is reason enough to reject it. A more technical objection is that it is hard to see how proto qualia could combine in such a manner as to "add up to" the sort of conscious experience we're familiar with in everyday life - an experience which seems to be a single conscious experience rather than a composite of billions of tiny proto-experiences, and which is present to a single conscious subject rather than to a collection of billions of tiny proto-subjects. A conscious experience, that is to say, has a unified character it would not have if it were an aggregate of simpler elements.

We will return later to the question of the unity of consciousness - a question which by no means poses a challenge to panpsychism alone. Its potentially panpsychist implications are, in any case, not the only problem for the Russellian theory. For it seems that the theory does not

#### 104 *Philosophy of Mind*

in fact avoid the zombie argument the way some of its defenders seem to think it does. Recall that what is essential to a molecule, atom, or subatomic particle qua molecule, atom, or subatomic particle is, in the Russellian view, that it plays a certain causal role, the role assigned to it in theoretical physics. The Russellian believes that qualia or proto qualia are what play these roles. But could something else have played them instead? There seems no reason not to think so. An analogy might help: what is essential to the particular philosophy professor Feser qua being a philosophy professor is that he is capable of teaching certain classes, directing students in their research, etc. Could someone other than Feser have performed those functions just as well? Much as he'd like to think otherwise, it is true that someone could. There is nothing about Feser qua Feser that is necessary to playing the role of being a philosophy professor: plenty of non-Fesers can and do play the role just as well. Similarly, there seems to be nothing about a quale or proto-quale qua proto-quale that is necessary to performing the functions of a basic physical particle. Something other than a proto-quale, something absolutely devoid of anything even vaguely analogous to qualitative character, could play the role just as well.

This would seem to entail that it really is perfectly possible for there to be a creature physical-particle-for-physical-particle identical to you which is utterly devoid of proto-qualia, and thus of qualia - a creature which has something other than proto-qualia playing the relevant causal roles. But then such a creature would be a zombie, in which case zombies really are conceivable even on the Russellian view. And if that is so, then even the Russellian view entails a kind of dualism: for it entails that qualia are one kind of thing, and the basic physical components of the universe qua physical (that is, quahaving the causal properties described by physical science), which can exist either with or without qualia, are another. Indeed, though Russell and Lockwood take themselves to be identity theorists of

a sort, Chalmers does not, and explicitly presents his own panpsychist brand of Russellianism as a version of property dualism.

Would a Russellian property dualism, like other forms of property dualism, be threatened with epiphenomenalism? At first glance, it might seem not: if qualia or proto-qualia are what play the causal roles physics associates with molecules, atoms, subatomic particles, etc., then they might indeed appear just obviously to have a causal influence on the physical world. But appearances are deceiving. Given that something other than proto-qualia could equally well play those same roles, there is nothing about their distinctly mental, qualitative character that

*Consciousness 105*

is relevant to their playing it. Feser is a husband and father, but his being a husband and father is completely irrelevant to his playing the role of a professor: someone who was neither a husband nor a father could play that role in exactly the same way. So Feser's being a husband and father is, we might say, epiphenomenal relative to his effects on the world qua philosophy professor. Similarly, a proto-qualia's qualitative character - being proto-reddish, or proto-pungent - is completely irrelevant to its playing the role of a subatomic particle: something lacking proto-reddishness or proto-pungency could have played the role in exactly the same way, so that these proto-qualitative features are epiphenomenal. So not only does the Russellian view lead to property dualism, but it seems to lead to epiphenomenalism too - with all the problems we've seen that entails.

### **A more consistent Russellianism**

Despite these problems Russell's theory might yet prove to be an advance over the usual alternatives. The reason lies not in the theory's metaphysical component - taking qualia to be the intrinsic properties of the material world, with all the weirdness this seems to lead to - but rather in its epistemology, its account of the nature of perceptual knowledge. Russell's central insight was, arguably, to see that indirect realism has dramatic implications for the mind-body problem; but it may have been an insight neither he nor his followers have taken seriously enough, or far enough.

Russell's own defense of indirect realism emphasized the causal element in perception, the way in which all our experiences of the external world are mediated by causal chains. The gap represented by these chains - by, for instance, the myriad neural firing patterns, retinal cell activity and stream of photons that come between the surface of an apple and your experience of it - entails, in his view, that you never directly get at external objects themselves, but at best only at mental representations of them. Russell assumed, however, that you do indeed, in introspection, directly get at these representations themselves. But do you?

In Russell's view, those perceptual representations are, like all other mental states, identical with certain brain processes, which come at the end of a long causal chain beginning with the surface of an external object. But then the introspection of these representations must be as dependent on the causal workings of the brain as perception is. If your perception of external objects is mediated by causal chains, surely so is

### *106 Philosophy of Mind*

your introspection of those perceptions, as brain events subserving perception, triggered by impulses from the sensory organs, in turn trigger further brain events subserving introspection. As with perception, introspection would thus seem to provide you with only a representation - an introspective representation - of what you are made aware of through it. It gives you a representation, that is to say, of your perceptual representations themselves; it does not acquaint you with the intrinsic nature of those representations. And if we imagine yet higher-order mental events directed on to introspection itself instances of meta-introspection, if you will - then these too must, on the Russellian model, be

regarded as involving yet further causal chains and thus yet higher-level representations (that is, representations of representations of representations).

If this is right; then there is reason to believe that we have, contrary to Russell, no more knowledge of the inner world of the brain as it is "in itself" than we have knowledge of the external physical world as it is in itself. All such knowledge would be mediated by representations. One consequence of this seems to be that the Russellian response to the zombie argument can be salvaged after all. Zombies really are inconceivable, for in conceiving of perceptual experiences and qualia as I encounter them in introspection existing apart from the abstract causal structure of the brain (or whatever), I am not conceiving of those experiences and qualia as they are in themselves, but only of introspective representations of them. As with Russell's original proposal, we can conclude that conceiving of that sort of thing existing apart from the brain is of no more consequence than is the fact that the symbol "H<sub>2</sub>O" can be imagined to exist in the absence of water. This would also appear to restore to the Russellian view its status as a version of neutral monism rather than property dualism. There is, at least where the question of the relationship between consciousness and the brain is concerned, only one kind of stuff, but it is intrinsically neither mental nor material. We count it as material when it is presented to us via perception, and as mental when presented to us via introspection: hence the brain seems "material" when one examines it during brain surgery, but "mental" when one "looks within" at thoughts, experiences, and feelings; but one is aware, of exactly the same object in both cases. The difference between material processes and qualia is a difference only in how we represent things, not a difference in the things themselves as they exist independently of us. It is, that is to say, an epistemological difference, not a metaphysical one.