

Intentionality

The term "intentionality" derives from the Latin *intendere*, which means "to point (at)" or "to aim (at)" - hence the use of the term to signify the capacity of a mental state to "point at," or to be about, or to mean, stand for, or represent, something beyond itself. (It is important to note that intentions, for example, your intention to read this chapter, are only one manifestation of intentionality; your belief that you are reading a book, your desire to read it, your Perception of the book, and so forth, exhibit intentionality just as much as your intention does.) The concept was of great interest to the medieval philosophers, but Franz Brentano (1838-1917) is the thinker most responsible for putting it at the forefront of contemporary philosophical discussion. Brentano's also famous for regarding intentionality as the "mark of the mental" - the one essential feature of all mental phenomena - and for holding that their possessing intentionality makes mental phenomena ultimately irreducible to, and inexplicable in terms of, physical phenomena. The previous two chapters gave us reason to think he was right to make the first claim. The present chapter will consider whether he was also right to make the second. In chapters 1 and 2 we examined some reasons for taking intentionality to be mysterious and perhaps incapable of a materialistic explanation. The intuitive idea was as follows: when we consider examples of material entities that exhibit intentionality - words, sentences,

pictures - we see that they do not have their intentional content inherently, but only relative to human interests; in itself, a word, sentence or picture is just a meaningless set of ink markings and has whatever meaning it has only because we use it to convey a meaning. As Searle has put it, the intentionality present here is "derived intentionality" rather than "intrinsic intentionality." (Searle also distinguishes a third category: "as-if intentionality," which something exhibits when it behaves as if it had intentionality though it really doesn't, for example, the way water in a river moves as if it wanted to get to the ocean, when in reality it doesn't "want" anything at all.) The derivativeness of their intentionality seems to be a necessary feature of the entities in question: since it is, intrinsically, just a collection of meaningless particles of ink, say, a written word or sentence couldn't have intrinsic intentionality. But what is true of these examples seems true of material entities in general. Sound-waves emitted by the larynx, electrical current passing through a computer and the like all have whatever intentionality they do only in a derived fashion. More to the point, brain processes, composed as they are of meaningless chemical components, seem as inherently devoid of intentionality as soundwaves or ink marks. Any intentionality they have would also have to be derived from something else. But if anything physical would be devoid of intrinsic intentionality, whatever does have intrinsic intentionality would thereby have to be non-physical. Since the mind is the source of the intentionality of physical entities like sentences and pictures, and doesn't get its intentionality from anything else (there's no one "using" our minds to convey meaning) it seems to follow that the mind has intrinsic intentionality and thus is non-physical.

In chapter 5 we considered the suggestion that the objection to identifying qualitative conscious states with brain states could be overcome by arguing, in modified Russellian "neutral monist" fashion, that neither perception nor introspection reveals to us the inherent nature of its objects: the way the brain appears to us in perception and the way conscious states appear to us in introspection are not necessarily the ways those things really are intrinsically. Perception and introspection give us only representations of the brain and of qualia-bearing conscious experiences, respectively, and not the real nature of those things as they are "in themselves." If it seems that brain states and conscious states cannot be identical, this might reflect just a difference in the way we represent them, and not an objective difference in the things themselves; they might, for all that, really be identical after all.

Could such a move be made in answer to the argument just sketched against identifying intentional mental states with brain processes? No, and the reason should be obvious. The modified Russellian neutral monist strategy depends on holding that the greyishness and squishiness of the brain are not intrinsic to the brain and that the qualia associated with conscious experiences are not intrinsic to the experiences: the greyishness, squishiness, and qualia all exist only relative to our representations of the brain and of conscious experiences. But the same move cannot be made with respect to the intentionality of intentional mental states. It would make no sense to hold that the intentionality exhibited by the mind does not exist intrinsically in the mind itself but only relative to our representations of the mind; for a representation is itself a manifestation of intentionality. We couldn't possibly "represent" ourselves as having intentionality unless we really had it, in which case we never have it only relative to a representation. That we can represent at all shows that we have it intrinsically.

Naturalistic theories of meaning

That, anyway, is the *prima facie* case for holding intentionality to be inexplicable in materialist terms. But despite this apparent difficulty for materialism - or perhaps precisely because of it - the attempt to provide a materialistic or "naturalistic" account of intentionality has been one of the main preoccupations of contemporary philosophers of mind. Some of them have suggested that the variety of accounts developed in recent years give, by virtue of their very existence, reason to think that a materialist explanation of intentionality should at least in principle be possible after all. So we need to consider these accounts and see whether they overcome the intuitive difficulties that seem to face such an explanation. The major theories can be grouped into four categories:

1. Conceptual role theories

This sort of theory proposes that the meaning or intentional content of any particular mental state (a belief, desire, or whatever) derives from the role it plays within a system of mental states, all of which, as we've seen, seem logically interrelated in the manner briefly discussed in chapters 3 and 6, since to have any one mental state seems to require having a number of others along with it. The idea is that what gives the belief that Socrates is mortal the precise meaning it has is that it is

entailed by other beliefs meaning that all men are mortal and that Socrates is a man, that together with a belief meaning that all mortals will eventually die it entails a belief meaning that Socrates will eventually die, and so on. If we think of beliefs, desires, and the like as a vast system of logically interconnected elements, the theory holds that each element in the system gets its meaning from having precisely the place in the system it has, by bearing exactly the logical and conceptual relations it bears to the other elements. (More precisely, it is the objects of beliefs, desires, and the like - sentences of Mentalese according to the CRTT, or, more generically and for those not necessarily committed to the CRTT, "mental representations" of some other, non-sentential sort - that bear meaning or intentional content. But for the sake of simplicity, we can ignore this qualification in what follows.)

There seems to be a serious problem with the conceptual role approach, namely that even if it is granted that mental states have the specific meaning or content they do only because of their relations to other mental states, this wouldn't explain how mental states have any meaning at all in the first place. That a particular belief either implies other beliefs or is implied by them presupposes that it has some meaning or other: nothing that was completely meaningless could imply (or be implied by) anything. The very having of logical and conceptual relations assumes the prior existence of meaning, so that no appeal to logical and conceptual connections can (fully) account for meaning. Moreover, if belief A gets its content from its relations to beliefs B and C, and these get their content from their relations to beliefs D, E, and F, we seem destined to be led either in a circle or to an infinite regress. Either way, no ultimate explanation of intentional content will have been given. To provide such an explanation thus inevitably requires an appeal to something outside the network, something which can impart meaning to the whole.

John Searle, who endorses something like the conceptual role theory of meaning, acknowledges that logical and conceptual relations between mental states cannot be the whole story if circularity or infinite regress is to be avoided. He therefore postulates that the entire "Network" of intentional mental states (he capitalizes Network to signify its status as a technical term) rests on what he calls a "Background" of non-intentional capacities to interact with the world around us. We have, for example, such intentional mental states as the desire to have a beer and the belief that there is beer in the refrigerator, and these mental states do, in part, get the specific meaning they have

via their relations to each other and to other mental states in the broader Network. But ultimately these mental states, and the Network as a whole, function only against a Background of capacities, such as the capacity to move about the world of physical objects, pick them up, manipulate them, and so on. This capacity is not to be identified with the belief that there is a real external world of physical objects; for if it were such an intentional mental state, then it would have to get its meaning from other mental states, and thus couldn't serve as part of the Background that ends the regress of mental states. The capacity in question is rather something unconscious and without intentionality, a way of acting rather than a way of thinking. One acts as if one had the belief in question, though one in fact does not. While this capacity could in principle become a conscious, intentional mental state - one could come to have the explicit belief that there is a real world of external physical objects that I can manipulate and move about within- this would mean that this particular capacity has moved out of the Background and into the Network, and now rests on some other unconscious, non-intentional Background capacity or way of acting. There is, in short, always some set of capacities or other that comprises the Background (even if it is not always the same set for different people, or even for the same person at different times), and these capacities serve to ground the Network of intentional mental states.

There is much to be said for Searle's hypothesis of the Background, but it seems that it cannot save the conceptual role theory, for to speak of a "non-intentional capacity for acting" is to speak ambiguously. Consider that when you act without the conscious belief that there is an external world of physical objects, but merely manifest a capacity to interact with the world of physical objects, your capacity isn't non intentional in the same sense that an electric fan's capacity to interact with the world of physical objects is non-intentional. You behave "as if" you had a conscious, intentional belief in a world of physical objects, but of course you don't, because it typically never even occurs to you either to believe or doubt that there is such a world: you just interact with the world, period. The fan also behaves "as if" it believed there was a world of external physical objects (that it "wants" to cool down, say); but of course it doesn't really have this belief (or any wants) at all. In the case of the fan, this is not because it just hasn't occurred to the fan to think about whether there is such a world, for the fan isn't capable of such thoughts; it is rather because, strictly speaking, the fan doesn't really "act" or "behave" at all, as opposed to

just making movements. And the reason we don't regard it as acting or behaving in the same sense we do is precisely because it doesn't have intentionality - it is a dumb, meaningless, unconscious hunk of steel and wires. We on the other hand don't merely make physical movements: the waving of your hand when your friend enters the room isn't just a meaningless movement, but an action, the action of greeting your friend. If it were just a meaningless movement – the result of a seizure, say - we wouldn't count it as an action at all; it wouldn't in that case be something you do, but rather something that happened to you. The fan, however, is capable of making nothing but meaningless movements.

For something genuinely to behave or act as we do requires that it does have intentionality - action and behavior of the sort we exhibit are themselves manifestations of intentionality, and thus presuppose it. But in that case, an appeal to a "capacity for action" cannot provide the ultimate explanation of intentionality. We need to know why our capacities for action are different from the mere capacities for movement that a fan exhibits. Merely noting, a la Searle's Background hypothesis, that our capacities are non-intentional ways of acting cannot help, for that they are genuinely ways of acting is precisely what needs to be explained. Indeed, since they are ways of acting, they cannot be literally non-intentional, for if they were, they would no more be true ways of acting than are the capacities of an electrical fan. A capacity for action is, as a matter of conceptual necessity, an intentional capacity.

In fairness to Searle, it isn't clear that he intends his hypothesis of the Background to serve as a complete explanation of intentionality. His aim may be just to draw out some implications of the fact that mental states are logically and conceptually related to one another in a Network. The point, though, is that his way of avoiding the circularity or regress that threaten any conceptual role theory cannot be appealed to in order to vindicate such a theory as a complete theory of meaning - and that it may even be incoherent, if Searle holds that the capacities and ways of acting that form the Background are literally devoid of intentionality.

2. Causal theories

The right way to break out of the circle or regress of mental states is, in the view of many contemporary philosophers, to appeal to the causal relations those states bear to elements of the external world. It is, in this view, not (or not merely) the relations these mental states bear to one

another that give them their intentional content or meaning, but (also) the fact that those mental states tend to be generated by certain kinds of interaction with the thinker's environment. Your belief that the cat is on the mat has the particular content it has not (merely) because of the logical and conceptual relations that belief bears to other mental states, but (also) because that belief tends to be caused by the presence of a cat in your external surroundings.

Some theorists would hold that causal relations alone account for the intentional content of mental states, while others would allow that conceptual role plays a part as well. The latter would accordingly distinguish between "wide content" (that aspect of the intentional content or meaning of a mental state that is determined by its causal relations to the external world) and "narrow content" (that aspect of intentional content or meaning that is determined by a mental state's relations to other mental states). Theories which, like causal accounts, tend to emphasize wide content are typically referred to as "externalist" theories (since they focus on causal relations to elements external to the thinker), while theories, like the conceptual role account, which tend to emphasize the priority of narrow content, are called "internalist" (since they focus on logical and conceptual relations between mental states, which are internal to the thinker). Externalist theories have in recent years come to be favored by philosophers of mind interested in giving a naturalistic account of intentionality.

The reason for this is not difficult to see. As noted earlier, the mind's evident causal interaction with the physical world provides the most powerful argument for the materialist claim that the mind must be just one more part of that world; and that the best way to account for the mind in materialistic terms is to analyze it into its causal relations is the central claim of the functionalism that has come to be the most popular version of materialism. It is natural, then, for the materialist to suspect that a causal approach to intentionality, in particular, is likely to succeed if any naturalistic approach will. Moreover, the causal approach clearly has some intuitive plausibility: surely, one is inclined to say the fact that your belief that the cat is on the mat was caused by the cat's being there has something to do with the fact that it has the content it does.

As it stands (and as all causal theorists recognize), the idea clearly needs development. For you could have the belief that the cat is on the mat even if that belief were not caused by the cat's being there, but instead caused by something else (like hallucinogenic drugs put into

your coffee), and the cat's being there could cause a belief other than the belief that the cat is on the mat (for example, because of bad lighting, it could instead cause the belief that the dog is on the newspaper). So mere causal connection is not enough to account for meaning. At the very least, some kind of regular correlation between a mental state and a particular cause of that state also seems crucial. Many philosophers see models for such correlations in the natural world: smoke is correlated with fire, the rings of a tree with its age, and the symptoms of a disease with the disease itself. So regular are these correlations that in each case we typically take the presence of the effect to provide a reliable indication of the presence of the cause: that is to say, we take the presence of smoke to be a reliable indication that fire is present, the presence of thirty-three tree rings reliably to indicate that the tree is thirty-three years old, and the presence of red spots to be a reliable indicator of measles. Indeed, we even use the language of meaning here: we say that smoke means fire, and so forth. Such regular, reliably indicative correlations seem a plausible model for the sort of causal connections that could explain the meaning or intentional content of mental states.

A little thought shows that even this development of the basic idea of a causal theory of meaning cannot be the end of the story, for how could even such regular causal correlations explain our ability to have thoughts about things we don't seem to have any causal connection with - non-existent objects (Superman and Santa Claus), future objects and events, and so on? Moreover, how can it explain our ability to make mistakes? In many cases a mental state "means" something with which it isn't causally correlated in a regular way: as we've seen, you might, because of bad lighting, take something to be a dog which is in fact a cat. Philosophers call this the "misrepresentation problem" for causal theories of meaning. A related, though distinct problem is the "disjunction problem": if (because there's always bad lighting in your house, or because you've got bad eyesight) a particular mental state of yours tends regularly to be caused not only by cats but also, in certain circumstances, by dogs, why (if the causal theory is true) should we regard that mental state as representing cats uniquely? Why should we not regard it as representing, disjunctively, cats OR dogs-in-certain-circumstances? Of course, there are going to be many cases where it does represent cats uniquely - the elderly person with bad eyesight might really only ever think that a cat is present, even when it's a dog - but that is precisely the problem: how can the causal theory explain this, given that the

theory seems to entail that your mental state will represent anything that regularly causes it?

Jerry Fodor, an influential proponent of the causal theory, has suggested that the solution to such problems lies in the notion of what he calls asymmetric dependence. The idea is that, when a mental state typically caused by cats is also caused by dogs-in-certain-circumstances, the latter sort of causal connection is parasitic on the first. That is to say, dogs-in-certain-circumstances will cause the relevant mental state only because that mental state is already typically caused by cats - the "dogs-in-certain-circumstances" causal connection only gets set up once the "cats" connection is in place - while cats would cause the relevant mental state whether or not dogs-in-certain-circumstances ever did. The causal connection between the mental state and dogs-in-certain-circumstances is therefore asymmetrically dependent on the causal connection between the same mental state and cats: the former connection will exist only if the latter does, but the latter would exist whether or not the former did. The right way to formulate a causal theory, then, is to hold that it is causal connections that are not thus asymmetrically dependent that give rise to meaning: in the case at hand, the mental state represents cats uniquely because the causal connection between it and cats is not asymmetrically dependent on some other causal connection.

Fodor's is but one attempt to solve the problems facing the causal theory, and all such efforts have faced a battery of further objections. The result has been the incorporation of ever more subtle and complex technical qualifications into the causal story in terms of which causal theorists want to account for meaning, so as to stave off various counter-examples. But even if one or more of these various technical moves can successfully deal with the specific counter-examples they are designed to handle, it seems that several fundamental difficulties facing any possible causal theory would remain unanswered.

The first problem is that the theory seems to assume that it is not possible for a mental state genuinely to represent anything other than something that typically causes it. But we have already seen, in chapter 1, powerful reasons for thinking that this assumption is false: your thoughts and perceptions might represent cats even if they are never caused by cats at all, but by a cartesian evil spirit, or a supercomputer stimulating your brain, as it sits in a vat of nutrients. A causal theorist might deny that this is really possible, but if so, he cannot appeal to the causal theory itself as grounds for this denial without begging the question.

A causal theorist willing to countenance the narrow content favored by the conceptual role theorist, in addition to the wide content emphasized by the causal theory, could perhaps reply that the possibility of these skeptical scenarios can be accounted for in terms of the former sort of content: the logical and conceptual connections one's thoughts about cats have to other mental states might suffice to make them genuinely about cats, despite the possibility that they are caused by something other than cats (for example, by a Cartesian evil spirit). It is not clear that this would work to save the causal theory - for if my thoughts would be thoughts about cats regardless of what caused them, how can causal relations play any role in generating meaning? But in any event, the causal theory would still remain open to an objection that we've already seen applied to the conceptual role theory. Like that theory, the causal theory would seem at most to account for why a particular mental state means this specifically, rather than that, it does not thereby account for why it has any meaning at all. It seems that it is only when a mind, with all its intentionality, has already come into being that there can be mental states which bear specific meanings related to their specific causes; and if so, then an appeal to such causes cannot by itself account for intentionality.

The causal theorist's appeal to alleged cases of meaning in nature does nothing to undermine the point: it supports it. When we say that "Smoke means fire," we're not speaking literally. Smoke doesn't really mean anything, at least not in the way that the word "smoke" means (given our linguistic practices) smoke. Smoke is just smoke - a meaningless arrangement of particles. Because smoke is typically caused by fire, we can interpret it as a sign of fire; but in that case the meaning is all in us, not in the smoke. That it's caused by fire explains why smoke "means" fire to us - that is, why it means fire rather than, say, water. But that it "means" anything at all has nothing to do with its causal connection to the fire and everything to do with our powers of interpretation and evaluation of evidence. In so far as the intuitive plausibility of causal theories of meaning rests on appeal to such examples as "smoke means fire," it thereby seems to rest on little more than a pun. The sense of "means" in that case just isn't the same sense as that in which your thought about fire means fire. The latter is a case of what Searle calls intrinsic intentionality, while the former is, when not a case of mere as-if intentionality at best a case of derived intentionality.

There is a third, and perhaps even deeper, objection to any possible causal theory. The point derives from an argument presented by Karl

Popper in the context of a critique of causal accounts of language, but it seems to be applicable to causal theories of intentionality as well'

Any account such theories could give of the relevant causal relations holding between a particular mental state and a particular object in the external world will require picking out a particular beginning point of the causal series (call it A) as the thing represented and a particular end point (B) as the mental state doing the representing. So suppose A is a particular cat you are looking at and B a particular brain state that the causal theorist wants to identify with the perceptual mental state representing the cat. The problem is this: in the external physical world as it is in itself, apart from human purposes and interests, there seems to be nothing more than an ongoing causal flux, comprising an unimaginably complex sequence of events. Nothing in this flux is objectively either the determinate starting point of a particular sequence of events or the determinate ending point. It is we who pick out certain events and count them as beginnings and endings; their status as beginnings and endings is relative to certain purposes and interests of ours. This is as true of A and B as of anything else: there is no objective reason why A should be the cat rather than the cat's fur or a particular photon in the stream leading from the cat to our retinas, and no objective reason why B should be this particular brain state rather than the one immediately before or after it in the causal sequence of brain processes. So the "fact" that the causal chain purportedly explaining your perceptual experience of the cat begins with A and ends with B would appear to be a mind-dependent fact, determined by human purposes and interests - that is to say, it appears to presuppose intentionality. But then, the characterization of all such causal chains would presuppose intentionality - in which case, no appeal to such causal chains could truly explain it after all.

3. Biological theories

Materialist philosophers of mind sensitive to the difficulties inherent in deriving meaning from brute causation have suggested that a more plausible candidate for a purely physical property capable of grounding intentionality might be found in the notion of biological function. Fins perform the function of allowing the organism having them to move through the water. Wings perform the function of allowing winged creatures to fly. Hearts perform the function of pumping blood. These organs serve these functions because natural selection formed them to do so. Might this sort of function underlie

the meaningfulness of mental states? It is, after all, surely the function of a desire to drink water to get a creature that has that desire actually to drink water, which a creature needs to do in order to survive and reproduce; that is plausibly why natural selection put such desires into creatures. And perhaps that's all it is for the desire to have the particular meaning or intentional content it has: its representing water is nothing more than its serving the function of getting the creature to drink water. Meaning, on this view, is identical to biological function -hence it is sometimes called a biosemantic theory of meaning (a label associated with Ruth Millikan, one of its main proponents).

An advantage of this theory is that it seems to provide a way of dealing with the misrepresentation problem. If the meaning or intentional content of a mental state derives from the biological function it serves, it will have that same meaning even if on some occasions it is caused by something other than what it is normally caused by. There thus need be no mystery about how a mental state could be about something other than what happens to cause it on some particular occasion, and thereby misrepresent what happens then to cause it. For example, if the desire to avoid snakes has the meaning it does because it serves the function of causing the creature having it to flee when snakes are present, it will still have this meaning even when a particular instance of it is caused, not by the presence of a snake, but by the presence of a rope or a hose that, due to odd lighting, looks like a snake.

Nevertheless, there are several serious objections to the biological theory. An obvious initial objection is that at best, it seems dubious that it could account for such sophisticated mental states as, say, one's belief that Wittgenstein was a more important philosopher than Russell: surely natural selection never hard-wired such a belief into anyone, for beliefs about the relative importance of Wittgenstein and Russell could not only not have occurred to anyone in the period of history in which natural selection formed human nature, but wouldn't have served any evident biological function even if they had occurred to anyone then.

Biosemantics advocates hold that such highly complex mental states might, nevertheless, derive a secondary functionality by virtue of their relationship to mental states - like the desire for water - that are more clearly functional. But however such a suggestion might be developed, there may be deeper problems. One of them is that the theory appears to entail that nothing that didn't evolve could possibly have intentionality for, not having evolved, it wouldn't have states that serve any particular function. But this seems false: we can certainly at least

imagine cases where creatures come into existence other than by evolution, and yet have intentionality. If a freak occurrence in a swamp were spontaneously to generate out of the muck a molecule for molecule living duplicate of you - "swampman," as philosophers who have discussed this sort of example have affectionately dubbed it – then this duplicate would surely have thoughts, experiences, and other intentional mental states, despite not having come about through evolution.

Another difficulty is that the biological theory seems unable to deal with the disjunction problem: if, for example, a desire to avoid cheetahs happened to be hard-wired into our ancestors as a result of their interactions with both cheetahs and tigers-in-certain-circumstances (for example, at night time when tigers might be hard to distinguish from cheetahs), then it would seem to follow that the biological function of this desire is to get us to avoid both cheetahs and tigers-in-certain-circumstances - and thus it would follow too that the desire represents, not cheetahs uniquely, but rather cheetahs OR tigers-in-certain-circumstances.

In reply to this, Daniel Dennett has suggested that if such examples indicate that meaning must be indeterminate on a biological theory of intentionality, this does not serve as an objection to the theory, for such indeterminacy is common throughout the biological realm. A certain organ may have evolved originally to serve one function, and then at a later stage in evolution taken on another: one creature might have evolved feathers because they served the function of attracting mates; while its descendants, having migrated to a colder environment, found that the feathers served to keep their bodies warm, a function the feathers might retain even if the mating function disappears. Which function the feathers really serve might, at some stage in this long evolutionary process, simply be indeterminate. But in that case, why couldn't the meaning of a desire to avoid cheetahs also be indeterminate (that is, not clearly about cheetahs uniquely as opposed to cheetahs OR tigers-in-certain-circumstances)? Why assume this is a problem for the biological theory, rather than just a further instance of the ambiguity evident in many biological phenomena?

One possible objection to this reply is that it fails to explain how the biological theory can deal with cases of mental states whose meaning or intentional content is determinate and unambiguous (the case for holding that there are indeed such mental states being something we'll examine shortly). Another objection is that even if all our mental states

were indeterminate or ambiguous in their meaning or content, this would not save the theory; for even if the theory could explain why they have ambiguous meanings, it would not explain why they have any meaning at all. While a heart serves the function of pumping blood, the heart nevertheless doesn't mean or represent pumping blood - for it doesn't mean or represent anything at all. It's just a muscle. Words, sentences, and pictures mean things, but muscles surely don't, any more than gall stones or hangnails do. But if having evolved to serve a certain function doesn't suffice to give the heart meaning or intentional content, why would this suffice to give a belief or desire meaning or intentional content? wouldn't mental states exhibiting intentionality already have to exist in the first place in order for natural selection to select some of them as having survival value? If so, then even if a mental state's serving a particular biological function could account for its having the specific meaning that it has (ambiguously or otherwise), it couldn't account for its having any meaning at all. Natural selection's purported ability to shape meaning would presuppose that there is meaning there to be shaped - in which case biological function couldn't possibly provide a full explanation of meaning.

This is, of course, an application to the biological theory of an objection already considered when discussing the conceptual role and causal theories - namely, that the operation of the mechanism the theory appeals to in order to explain intentionality itself presupposes intentionality. That this criticism seems to apply to the biological theory as much as to the causal theory is even more evident when one considers that ultimately, there may be no substantive difference between them. For, as Searle has argued, the trouble with appeals to biological function in this context is that all talk about biological function must, from a Darwinian point of view anyway, be regarded as nothing more than a shorthand for talk about causation. To say that the heart was selected by evolution to serve the function of pumping blood is, strictly speaking, to say something false; for evolution doesn't literally "select" anything, nor does the heart literally serve any purpose or function at all, at least not on a Darwinian view. Indeed, the whole point of Darwin's account of evolution by natural selection is to get rid of the need to appeal to literal purposes and functions in nature - to explain the appearance of purpose and function in terms that make reference only to purposeless, meaningless causal processes. The right thing to say about the heart is, in a Darwinian view, just this: it causes blood to flow and it was in turn caused by a series of successive genetic

mutations that allowed the creatures exhibiting them to survive and reproduce in greater numbers than those which lacked them. And that's it. If talk about the "purpose" or "function" for which the heart was "selected" has any application at all, it is only as a way of noting how what in reality are the purposeless, functionless, and meaningless results of unthinking causal processes can seem to be purposive, functional, and meaningful.

Talk about purposes and functions, if taken literally, seems to presuppose intentionality; in particular, it seems to presuppose the agency of an intelligence who designs something for a particular purpose or to serve a particular function. But the aim of Darwinian evolutionary theory is to explain biological phenomena in a manner that involves no appeal to intelligent design. As we've had reason to note in earlier chapters, just as modern physics has tended to explain phenomena by carving off the subjective qualitative appearances of things and relocating them into the mind, so too did the Darwinian revolution in biology push purpose and function out of the biological realm, making them out to be mind-dependent and devoid of objective reality. This is of a piece with the general materialistic tendency to regard genuine scientific explanation as requiring the stripping away of anything that smacks of the subjective, first-person, intentional point of view. It thus seems odd that materialist philosophers should think it a hopeful strategy to appeal to biological function in order to account for intentionality. As Searle argues, this move is simply not open to them, given what is entailed by a Darwinian account of the biological realm - an account materialists must necessarily be deeply committed to.

4. Instrumentalist theories

Though Dennett, as indicated, sympathizes with biological theories of intentionality, he has also developed a distinctive approach of his own. It begins by proposing that what we're trying to understand in explaining intentionality is the behavior of certain complex physical systems: human beings and, perhaps, other animals. In explaining the behavior of a physical system, Dennett says, we can take one of three different "stances" toward it. We can, first of all, take what he calls the physical stance toward it, accounting for its behavior in terms of the laws of physics and the other natural sciences. This is the stance we typically take toward simple physical phenomena, whether in everyday life or in science. If we're trying to predict the course of a billiard ball or the consequences of mixing certain chemicals, it usually suffices to

think of these phenomena as governed by basic physical laws. Sometimes, however, the physical stance is unhelpful. If we're trying to understand the workings of a bodily organ - the heart, say - or of a machine - an automobile, perhaps - then we won't get very far by treating these things merely as physical systems governed by basic scientific laws. Confidently predicting and explaining the behavior of such systems is made possible by adopting instead the design stance toward them, which involves considering them as performing a certain function. To think of the heart as a collection of basic particles governed by the laws of physics isn't going to help you diagnose arrhythmia, but thinking of it instead as an organ whose function is to pump blood will. There are cases, however, where even the design stance is insufficient to allow us to explain and predict a system's behavior. The fact that a chess-playing computer was designed to serve the function of playing chess won't help us to guess what its strategy against Kasparov will be; that we take a mouse to be "designed" by natural selection to avoid predators won't tell us what path it will take in order to escape an oncoming cat. Here, Dennett says, we need to take the intentional stance, which involves regarding something as all "intentional system" - an entity having beliefs, desires, and other mental states and being capable of reasoning on the basis of them - and predicting and explaining its behavior accordingly. We say that the mouse thinks there is a place to hide over there or that the computer intends to employ the French defense, and are thereby enabled to understand what the mouse or computer does.

Which of these stances is the correct one to take in a particular situation? It depends. If you're trying to determine, not what moves the computer will make, but how many people it will take to lift it and carry it over to the room where the match will be held, the physical stance rather than the intentional stance will be most appropriate; if you're trying to figure out how to turn the computer on, the design stance will be the one to take. In each case, the right answer depends only in part on objective features of the system itself; it depends also, and ultimately, on our interests. Does a computer really have the intentionality we attribute to it in taking the intentional stance toward it? Dennett's answer is that if it is complex enough in its behavior that we cannot usefully explain and predict that behavior without taking the intentional stance, then it has all the intentionality a thing could possibly have. But its having it is, again, ultimately a function of our finding it useful to treat it as having it. And all intentionality is in Dennett's view like this, including our own: we regard ourselves as

intentional systems because that is the most practical way of dealing with ourselves and each other, of explaining and predicting our behavior. There is nothing more to it than that.

Given that Dennett seems not to be a realist about intentionality (that is, someone who takes it to exist independently of our purposes and interests) nor even, as causal and biological theorists appear to be, a reductionist who wants to reduce intentionality to something more basic (causal relations or biological functions), he is often classified instead as an instrumentalist- as holding that talk about the intentionality of our minds is a useful instrument or tool for understanding our behavior, but doesn't describe anything that exists objectively, independently of our purposes and interests. Thus understood, his view seems open to an obvious and seemingly fatal objection: for us to take a stance toward something, including the intentional stance, is itself a manifestation of intentionality; so we can't coherently suppose that intentionality is a mere artefact of the stance we take toward ourselves.

Dennett's reply to such an objection would seem to lie in his influential strategy of homuncular decomposition. The idea is this. We can usefully regard our minds as comprised of a number of subsystems that perform various mental functions: visual processing, linguistic competence, and so on. Each subsystem can itself be metaphorically understood as a "homunculus" - a "little man" who performs some particular task. But the functions performed by each of these homunculi can, like our own minds, be thought of as comprised of yet more basic functions performed by smaller subsystems; in other words, each of the homunculi comprising our own minds can be thought of as comprising smaller homunculi of its own. At the level of our minds as a whole, we are dealing with what we have reason to treat as systems possessing a very high degree of intentionality. But the homunculi that comprise our minds, precisely because they perform more specific, less comprehensive functions, possess a lower degree of intentionality; and by the same token, the smaller homunculi that comprise them possess even less intentionality. If we keep decomposing each level of homunculi into ever smaller levels, eventually we will come to a basic level of homunculi who, because they perform functions as simple as possible, have as little intentionality as possible. Think of these as extremely stupid homunculi - homunculi whose task is no more complicated than flipping a switch back and forth.

Such a task could, of course, be performed by a very simple machine. Yet it is not at all implausible to suppose that whatever

intentionality was possessed by such a machine would be intentionality that existed only relative to the stance we might take toward it. But then it shouldn't be implausible to suppose that the intentionality possessed by the very stupid homunculi that comprise the most elementary level of the subsystems comprising our minds should be explicable in terms of the intentional stance - in which case the intentionality possessed by our minds as a whole, which is just a composite of the intentionality possessed by its various subsystems, is also so explicable.

This strategy is not without ingenuity, but that it fails genuinely to answer the objection at hand seems to follow if we accept Searle's distinction between intrinsic intentionality on the one hand, and derived and as-if intentionality on the other. Machines, of course, have whatever intentionality they have only in either a derived or an as-if way. But our intentionality is intrinsic. So if there really are basic homunculi comprising our minds, their intentionality too must be intrinsic - in which case they are not comparable to machines, which do not have intrinsic intentionality. The intuitive force of Dennett's argument seems to rest on his comparison of the stupidity of the basic homunculi and the stupidity of a machine. But the two are not "stupid" in the same sense. The homunculi are stupid because they have extremely low intelligence; the machine is stupid because it has no intelligence at all. Strictly speaking, the machine isn't really even stupid in the first place, because one has to have at least a very minimal level of intelligence even to count as stupid (by comparison, that is, with those with higher intelligence). The machine doesn't even rise to the level of stupidity while the homunculi have at least that much going for them.

If their intentionality is intrinsic, then it isn't merely an artefact of our taking the intentional stance toward ourselves. Dennett might deny that it really is intrinsic - he might hold that the homunculi, no less than the machine, have at most derived or as-if intentionality. But if he says that the intentionality of the homunculi - and thus of our minds as a whole - is derived, he's back in the incoherent position of saying that we have intentionality only because we take ourselves to have it (where "taking ourselves to have it" is itself a manifestation of intentionality). And if instead he says that our intentionality is only "as-if," then he's saying something even more radical: that our intentionality doesn't really exist at all. But that brings us to another theory.

Eliminativism again

If the upshot of our discussion thus far seems to be that no naturalistic account of intentionality has yet succeeded, there are a number of materialists who would nevertheless deny that this has any tendency to cast serious doubt on the truth of materialism. What it really casts doubt on, they suggest, is the reality of intentionality itself. Recall from chapter 5 that some materialists have proposed that the way to deal with the problem posed by qualia is simply to deny that qualia exist in the first place. Many of them would apply the same strategy to the problem of intentionality: if intentional mental states turn out to be irreducible to purely material states of the brain, so much the worse for intentional mental states. We ought to stop looking for a way to reduce them, and instead consider eliminating them from our ontology altogether. Maybe they don't really exist at all, in which case there's no need to explain them.

This is the view known as eliminative materialism, most famously associated with Patricia and Paul Churchland, and if it doesn't sound utterly bizarre, you haven't understood it. Nor is this a biased description of the theory: eliminativists are under no illusions about how counter-intuitive and contrary to common sense their position is. They are willing frankly to deny what the average Person would consider undeniable, namely that we have thoughts, experiences, beliefs, desires and all the rest - in short, that we have minds. The eliminative materialist view is, not that mental states are identical to brain states or that minds are identical to brains, but rather that there are no mental states, and in short no minds, at all. There is only the brain, and whatever a completed neuroscience will eventually tell us about it. The correct description of human nature will, at the end of the day, make no reference to what we think, feel, hope, fear, or believe, but instead only to physiological structure, neuronal firing patterns, chemical secretions and the like. It's not that your belief that it's raining is the same thing as such-and-such a neural process, as the identity theory would have it; it's rather that neither you nor anyone else has ever had any beliefs, nor any other mental states at all, and that the neural process is all there is and all there ever has been.

Why would anyone take such a proposal seriously? Part of the answer has to do with the notion, discussed in chapter 3, that our commonsense description of ourselves as having beliefs, desires, and other mental states constitutes a kind of theory: "folk psychology." There we noted that the identity theory can be understood as

suggesting that this theory can be reduced to some neuroscientific theory in just the way that the theory that made reference to genes was shown to be reducible to a theory making reference to DNA. But, as the Churchlands are fond of noting, there are cases in the history of science where a theory turns out not to be reducible to some deeper theory, but instead to have been utterly mistaken and thus in need of elimination. The pre-Copernican picture of the universe, according to which the earth was at the center of the solar system and surrounded by a series of heavenly spheres, was just wrong: it was not reduced to modern astronomy, but eliminated and replaced by modern astronomy. And if folk psychology is a theory, then it too might turn out to be mistaken. Moreover, since, unlike other scientific theories, it seems not to be reducible to some more basic theory that makes reference to nothing but purely physical laws and entities, this is itself a reason to think it might be false. We are better off just getting rid of it, and reconceiving of human nature entirely in terms of the purely materialistic categories of neuroscience. This might not be possible immediately - we still have much to learn about the brain and nervous system - but, at least in principle, and at some future date, we ought to be able to substitute a wholly neuroscientific description of ourselves for our current mentalistic idiom. Perhaps the citizens of eliminative materialist societies of the future will no longer say things like "Boy, this pain is really getting to me," but rather "there's a particularly high level of activity in my C-fibers and reticular formation." Someone getting off an amusement park ride will no longer report feeling dizzy, but instead note that "there's a residual circulation of the inertial fluid in the semi-circular canals of my inner ear." Romance novelists will eschew talk of love and longing in favor of neuronal action potentials and behavioral dispositions.

All this may seem pretty fanciful, but that doesn't prove it is false. As eliminativists never tire of pointing out, they all laughed at Jules Verne too, until Neil Armstrong set foot on the moon. But there may be much deeper problems for eliminativists than merely being ahead of their time. First, the notion that folk psychology is a kind of theory seems much less plausible than the view that regards it instead as a description of the data that any theory worth our attention must be consistent with. (And if the indirect realist view discussed in chapter I is correct, then it is precisely our direct awareness of mental phenomena that forms the starting point of all our theorizing about the mind and its relationship to the physical world, so that it can hardly make sense to suggest that such phenomena do not exist.)

There seems to be an even more basic, and more obvious, difficulty with the theory however. In so far as eliminative materialism asks us to reconceive human nature, to learn more about the nervous system, and indeed to believe the theory itself, doesn't it presuppose the validity of the very concepts it proposes eliminating? Doesn't the theory ultimately contradict itself?

Eliminativists are, of course, well aware of this objection, but think it can be easily answered. The Churchlands propose that to accuse the eliminativist of self-contradiction is like accusing modern biologists of contradicting themselves in denying the pseudo-scientific concept of "vital spirit." It would obviously be foolish for vitalists to argue that people who disbelieve in vital spirit must be alive in order even to express their disbelief, in which case they must possess vital spirit after all and have thereby refuted themselves. Vitalists would, in this case, be begging the question, since their argument would presuppose that the only way to explain life is in terms of vital spirit, which is exactly what the anti-vitalist denies. Similarly, the reply goes, the critic of eliminative materialism is begging the question in assuming that eliminativists must "believe" their own theory, etc., since the existence of beliefs is exactly what eliminativists reject.

One reason to suspect that this reply will not work is that here again the eliminativist seems, unavoidably, to make use of concepts - "begging the question," "assuming," and the like - that are just the kind of mentalistic notions eliminative materialism denies the legitimacy of. This suggests that the analogy with vitalism may not be a good one. Anti-vitalists don't deny the existence of life, after all; they only deny a certain theory about how to explain life. That's why they aren't contradicting themselves, which they would be doing if they denied the existence, not only of vital spirit, but of living things (including themselves). But eliminativists don't just deny a certain theory about how to explain believing, assuming, thinking, etc.; they deny the very existence of these phenomena. Yet "denying" is itself an instance of the sort of phenomena whose existence is denied. In short, any attempt either to propose or reject a theory - eliminative materialism, folk psychology, or whatever - is to represent the world as being a certain way, and thereby to manifest intentionality. But in that case one cannot coherently propose a theory that denies the existence of representing or of intentionality.

Some eliminativists would acknowledge that their position has a real difficulty here, but suggest that it may be that we just don't yet have the conceptual resources to imagine how theories might be

proposed, accepted, and rejected without using mentalistic and intentional language. We might be in a position similar to someone in ancient Greece trying to imagine quantum mechanics: the theoretical groundwork necessary even to conceive of the radically novel conceptual scheme eventually to be developed just hasn't been laid yet. Again, though, even to frame this suggestion the eliminative materialist has to use language - "imagine," "conceive," "conceptual resources," even "theory," "propose," and "reject" - which seems irreducibly mentalistic and intentional. Anything that could ever count as a "theory," or even as something relevantly analogous to a theory - no matter how far off in the future it is put forward - would seem unavoidably to be something that involved representation and intentionality, in which case it could not coherently be used to express eliminative materialism. It is as if the eliminativist were suggesting that $2 + 2 = 23$ and that the only reason we can't make sense of this is that we don't yet have the conceptual resources to see what addition might look like in the future. The right thing to say about this is that whatever the people of the future might be doing if they go around asserting that $2 + 2 = 23$, it won't be addition of any sort. Similarly, eliminative materialists seem ultimately to be proposing a theory which is by their own admission currently unintelligible, with a promissory note that someday we might be able to make it intelligible. But the promise can in principle never be kept, since the possibility of intelligibility - which requires that we be able to understand or make sense of something, and thus involves intentionality - is something the theory itself appears to rule out as impossible.

Eliminativists seem in effect to be inviting us not to believe them or their theory, now or ever. So how can they blame anyone who takes them up on their offer?

The Indeterminacy of the physical

So far we have seen that all extant materialist attempts to deal with intentionality appear to face serious difficulties. The intuitive anti-materialist argument from intentionality with which we began this chapter remains, as yet, undefeated. But things are, in the view of some critics of materialism, even worse for the materialist than so far indicated. In their view there is, in addition to the fact that nothing material would seem capable of having any intrinsic meaning, a further problem: even if something physical could have intrinsic meaning, it could not by itself have the determinate meaning that (at

least many) mental states have. The argument is, in short: at least some intentional mental states and processes are determinate in their meaning; no physical state or process could possibly be determinate in meaning; so intentional mental states cannot be identified with or reduced to physical states and processes.

We've already seen a number of ways in which physical processes can be inherently indeterminate: for instance, we noted in the previous chapter that it is indeterminate from its physical properties alone what interpretation is to be assigned to the algorithmic rules governing a computer, and it was suggested earlier in this chapter that it is indeterminate from the physical facts alone what counts as the beginning or end of a given causal chain. It is to a generalization of considerations like these that some critics of materialism have appealed in developing not just an objection to this or that specific materialist theory, but a comprehensive anti-materialist argument from the determinacy of meaning. If such an argument were to succeed, it would have the effect of buttressing Searle's suggestion, considered at the end of the previous chapter, that it is indeterminate from the third-person behavioral and neurophysiological facts-alone what meaning is to be assigned to a person's utterances. The upshot of Searle's argument was that meaning must inevitably be determined from the subjective first-person point of view, and this, together with the considerations we want now to examine, arguably tends to reinforce the suggestion, considered over the course of the previous few chapters, that first-person, subjective facts are inexplicable in terms of third-person objective physical facts. These considerations will also suggest a way of responding to the possible objection to Searle considered at the end of the previous chapter, to the effect that mental states themselves may not be any more determinate than are physical processes.

The considerations in question concern three inter-related manifestations of intentionality: our use of representations, our grasp of concepts, and our capacity for formal reasoning. Let us consider each of these.

1. Representations:

We've already taken note of the concept of "mental representations," and sentences of Mentalese as possible candidates of what form mental representations might take. But let us now consider for a moment a much more pedestrian example of a representation - a drawing you might make of your mother. When you draw your mother, you are

creating a kind of representation of her. But notice that it is not the particular physical features of the drawing itself - the form of the lines you make, the chemicals in the ink you use, and so forth - which make it a representation of her. The reason is not merely the one noted at the beginning of the chapter, namely, that nothing physical seems capable of having any intrinsic intentionality; the reason is rather one that would apply even if the argument of the beginning of the chapter were to be rejected. Someone looking over your shoulder as you draw might later on produce an exact copy of the drawing you were making. Perhaps the person admires your craftsmanship and wants to see if he or she can do as well. But in doing so the person would not, strictly speaking, be drawing a representation of your mother - he or she may have no idea, nor any interest in, who it was that you were drawing but rather a representation of your representation. And, in general, the very same image could count either as a drawing of an X, or as a drawing of a drawing of an X - or indeed (supposing there's someone looking over the shoulder of the second artist and copying what he or she was drawing) as a drawing of a drawing of a drawing of an X, and so on ad infinitum.

Even if we count something as a drawing, and therefore as possessing some intentionality or other, exactly what it is a drawing of is still indeterminate from its physical properties alone. The same is true not just of drawings, but also of written and spoken words (for to say or write "cat" could be to represent cats, but it could also be to represent the word "cat") and indeed of any material representation, including purported representations encoded in neural firing patterns in the brain. There seems in general to be nothing about the physical properties of a material representation that make it a material representation of an X as opposed to a material representation of a material representation of an X.

Sometimes, however, you are determinately thinking about a particular thing or person, such as your mother. Your thought about your mother is about your mother - it represents your mother, and doesn't represent a representation of your mother (representations, pictures, and the like might be the furthest thing from your mind). But then your thought, whatever it is, cannot be entirely material. Given that there's nothing about a material representation per se that could make it a representation of an X as opposed to a representation of a representation of an X, if your thought was entirely material then there would be no fact of the matter about whether your thought represented your mother as opposed to a representation of your mother.

Your thought is determinate; purely material representations are not so your thought is not purely material.

The materialist might reply that we shouldn't look at a material representation in isolation to determine what it represents, but ought also to consider factors such as its conceptual relationships to other representations, its causal relations to the world, and the behavioral dispositions of, or rules followed by, the thinker having the representation. But as we've seen, such appeals to conceptual role, causation, behavioral dispositions and rules have serious problems of their own, some of which also concern indeterminacy. In particular, if the suggestion is that some system of material representations, of causal relations, behavioral dispositions, rules, or whatever, determines meaning, the problem is that the same difficulties which arise when we consider a single representation in isolation just recur at a higher level. Any such system of material elements or principles is indeterminate between alternative interpretations: but our representations seem, at least sometimes, not to be indeterminate in this way.

2. Concepts

In thinking about something, we bring it under a concept: we think of it as a cat or as a mother. And of course, we can also bring it under more than one concept - we might think of the same creature as both a cat and a mother. Either way, there is typically some determinate concept or concepts under which we bring whatever we are thinking about. We think of something as a cat or as a cat and a mother, say, and not as a dog or a father.

What is it that determines that our thoughts involve bringing something under exactly this particular concept or set of concepts rather than some other one? It seems it cannot be material facts alone. For instance, it cannot be some sort of physical relation (for example, a causal relation) of ours to all actual cats that makes our thoughts about cats involve applying the concept cat to them. For, to borrow an example from John Haldane, it could be that all actual cats also fall under a concept we can call maxifourn, where maxifourns are the most-common-four-legged-animals whose-average-weight-is-W. Any physical relation of ours to cats will therefore also be a relation to maxifourns. But our thoughts about cats nevertheless involve applying the concept cat, and do not involve the concept maxifourn. In that case, it cannot be the physical relation alone that determines what concept we're applying.

As Haldane notes, the point is even clearer with examples like triangle and trilateral, which are concepts applying to exactly the same

objects in every possible world (unlike cat and maxifourn, since these concepts will not apply to the same things in possible worlds where it is dogs rather than cats who are maxifourns). No physical relations between us and such objects can be sufficient to determine that we are thinking of them as triangular rather than trilateral. In general, there are always more ways to conceive of the objects of our thoughts than the physical facts can determine.

Related to this point is the consideration that concepts are inherently abstract and universal, whereas material phenomena are concrete and particular. accordingly, a concept cannot be identified with anything concrete, particular, or material; and thus it cannot be identified with any physical symbol in the brain or nervous system. Nor can it plausibly be identified with a set of behavioral dispositions, as is sometimes suggested, since, as noted above, behavioral dispositions are susceptible of various interpretations and are thus indeterminate in a way that (at least many) concepts are not. For similar reasons, the propositions we grasp, assent to, and deny – and of which concepts are the constituents - cannot be identified with "sentences in the head" or with any other material entities. Propositions are necessarily abstract. Had there been no human beings, the proposition there are no human beings would have been true, even though there would then have been no "sentence in the head" for that proposition to be identical to. Had there been no physical world at all, the proposition there is no physical world would have been true, even though there would then have been no physical entity of any sort for that proposition to be identical to. Some propositions are necessarily true, that is, true in all possible worlds, but no physical entity exists in all possible worlds (for example, there are possible worlds where there are no brains, and thus no "sentences in the head").And so on.

This obviously poses a further problem for the Mentalese hypothesis, and for any theory which takes thought to consist of nothing but material processes. In the view of some critics of materialism, it also suggests a further general anti-materialist argument: when the mind grasps a concept or proposition, there is clearly a sense in which that concept or proposition is in the mind; but if these things are in the mind and yet (for the reasons given above) cannot be in the brain, it would seem to follow that the mind cannot be identified with the brain, or for that matter with anything material.

3. Formal reasoning

Whatever one thinks of such an argument (and it surely stands in need of further development), the topic of abstract thought brings us to one last respect in which mental states, especially thoughts concerning necessary truths, can be determinate in a way material processes are not. When we make judgments of a mathematical or logical sort, our judgments have a certain determinate form: the form of addition or squaring, for instance, or of modus ponens, conjunction, or disjunction. Nothing that does not have exactly the form of $2 + 2 = 4$ counts as adding 2 and 2 to get 4; nothing that does not have exactly the form that If Socrates is a man, then Socrates is mortal; Socrates is a man; so Socrates is mortal has counts as an instance of reasoning via modus ponens. But, as James F. Ross has argued, no physical process can have the determinate form had by such formal thought process. Just as a paper plate or a Frisbee can approximate a "perfect" circle but can never truly realize one - that is, Paper plates, Frisbees and all other "circular" physical objects are never really circles at all, Strictly speaking (every true circle is already a "perfect" circle) - neither can any physical process ever do more than approximate formal reasoning.

When one considers the circle analogy, the intuitive plausibility of this claim already becomes evident. But Ross appeals to a number of results in recent philosophy to bolster the argument. One of them is Quine's argument for the indeterminacy of translation, already considered in the previous chapter. Quine argues that if the physical facts about us are all the facts there are, then there is no fact of the matter about what any of our utterances mean: meaning will be indeterminate. This would entail that our reasoning processes would also be indeterminate; there would be no fact of the matter about whether we are applying modus ponens or only some approximation to it. Another-relevant example is Saul Kripke's distinction (which Ross adapts for his own purposes) between addition and what he calls "quaddition," where addition has the form " $x + y$," but quaddition has the form " $x + y$, if $x, y < 57$, = 5 otherwise" A calculating machine doing addition and a machine doing quaddition will give the same results when the numbers they are computing are less than 57, but when the one doing addition computes 58 and 100 it will get 158, whereas the one doing quaddition will instead get 5. Because they'll get the same results in the first case, there is no fact about their behavior that can then determine whether they are doing addition or

quaddition. But suppose the difference in the results would manifest itself, not at 57 but instead at some much higher number - indeed, at a number that is higher than the highest number either calculator is capable of displaying. Then there would be, not only no way of knowing which of either of the machines was doing quaddition instead of addition, but no fact of the matter at all about which was performing which. The physical facts about the calculators are equally consistent with either addition or quaddition, and thus indeterminate between them. But if, as with calculating machines, the physical facts about us were all the facts there are, then it would be indeterminate with us too whether we are performing addition or quaddition. But it isn't indeterminate: we do addition, period. Our doing so thus cannot be a purely material process.

Some materialists - Quine, and perhaps Dennett - might reply that the right conclusion to draw from all this is that since (they claim) we are purely material beings, we just don't in fact add, or do modus ponens, or carry out any other piece of formal reasoning after all; it only seems like we do because we approximate doing so. In fact, they might say, all thought is as indeterminate as physical Processes are. However, this move is not only highly counterintuitive - it entails that you've never once added 2 and 2 to get 4, for example, but only think you have - but it threatens every argument that anyone has ever given, including every argument anyone has ever given for materialism. For if none of us ever really reasons via modus ponens or any other valid argument form, then we never reason validly. Every single argument anyone has ever given will have been invalid! This materialist response would thus undermine itself.

This shows just how extreme and costly is the suggested reply to Searle considered at the end of the previous chapter. It also indicates that such a reply cannot succeed, for the claim that none of our thoughts is determinate seems demonstrably false. As Ross notes, even to deny that we really have determinate thoughts, certainly where the thoughts in question concern addition, modus ponens, and the like, presupposes that we have determinate thoughts; for even to deny that we ever add or do modus ponens requires that we grasp these operations, and to grasp them is to have a thought of a form as determinate as that which is grasped.

materialism, meaning, and metaphysics

Arguments of the sort considered in the previous section go back as far as Plato and Aristotle; indeed, their contemporary proponents typi-

cally present them as merely reformulations in modern guise of essentially Platonic or Aristotelian lines of thought. In so far as these arguments tend, in essence, to expand some of the objections made to specific recent materialist theories of intentionality into comprehensive critiques of materialism, they illustrate the point I made in chapter 3 that many of the criticisms directed at materialism today are but variations on the same objections that have been made for two and a half millennia. This point is further bolstered if we accept the intentionalist thesis that to be a subject of conscious experience is just to be a subject of certain intentional states, so that the problems of consciousness and qualia - often thought to constitute distinctively modern challenges to materialism - really boil down at the end of the day to the ancient problem of intentionality.

In summary, the difficulty intentionality seems to pose for the materialist is this: if Searle is right, intrinsic meaning or intentionality and the first-person point of view of the conscious, thinking subject are inextricably bound up together; and if the arguments of the preceding section are right, meaning or intentionality, and thus the first-person point of view of the conscious, thinking subject, are irreducible to and inexplicable in terms of anything material, including the brain. Dualism would seem to be vindicated.

Materialists might, nevertheless, suggest that we shouldn't be too quick to draw such a conclusion. For is the dualist really in any better a position than the materialist where meaning or intentionality is concerned? How, after all, does appeal to the existence of a nonphysical subject or non-physical properties explain intentionality? Hasn't the dualist really just supplemented one mystery - the mystery of intentionality - with another, namely the mystery of the nature and operation of non-physical minds? And doesn't the interaction problem that has plagued the dualist since Descartes's time indicate that this second mystery is itself unlikely to be solved?

As we've seen when considering the argument from reason and the problem of mental causation, it isn't quite right to say that the interaction problem poses a challenge to the dualist alone, but the questions just raised are fair. The dualist might respond that the point of arguments of the sort considered in the last section isn't to explain intentionality in the first place but rather to demonstrate that whatever intentionality is, it isn't physical. And if it isn't, to try to find some physical explanation of it will be a waste of time. Of course, the materialist might complain that non-physical processes are not the kind that can possibly be studied via the methods of physical science. But to this

the dualist could reply that it is a mistake to think that physical science is the only legitimate avenue of inquiry. The proper approach to the study of the mind, in the dualist's view, is via metaphysics rather than physics, and philosophy rather than natural science. For since, in the dualist's view, the arguments for dualism show that the mind is nonphysical, they thereby show also that it is only via inquiry other than scientific inquiry that we are going to understand its nature, if we are going to understand it at all. For the materialist to reject the possibility of such inquiry, a priori, would simply be to beg the question against the dualist.

But can metaphysics really say anything to clarify the nature of non-physical minds that hasn't been said already by Descartes and his successors? That brings us to the topic of our final chapter, where we will see that, as with some of the arguments we've considered in this chapter, dualists may be well advised to look to their ancient rather than modern forebears to find the most promising means of defending their position.

Further reading

Brentano's famous analysis of intentionality is to be found in his *psychology from an Empirical Standpoint* (London: Routledge and Kegan Paul, 1973). Interest in the topic was renewed among analytic philosophers by a famous exchange between Wilfrid Sellars and Roderick Chisholm entitled "Intentionality and the Mental," in Herbert Feigl, Michael Scriven, and Grover Maxwell, eds., *Minnesota Studies in the Philosophy of Science, Vol. II: Concepts, Theories, and the Mind-Body Problem* (Minneapolis: University of Minnesota Press, 1958). John Searle's distinction between kinds of intentionality and critique of the appeal by naturalistic theories of intentionality to the notion of biological function are to be found in his *The Rediscovery of the Mind* (Cambridge, MA: The MIT Press, 1992). His own account of intentionality, including the role of the Network and the Background, is developed in his *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press, 1983).

An excellent source of articles representing all the main naturalistic theories of intentionality is Stephen P. Stich and Ted A. Warfield, eds., *Mental Representation: A Reader* (Oxford: Blackwell, 1994). Some of the more influential full-length studies are: Fred Dretske's *Knowledge and the Flow of Information* (Cambridge, MA: The MIT Press, 1981), Jerry Fodor's *Psychosemantics: The Problem of Meaning*

in the *Philosophy of Mind* (Cambridge, MA: The MIT press, 1987), and Fodor's *A Theory of Content and Other Essays* (Cambridge, MA: The MIT Press, 1990), all of which defend versions of the causal theory; Ruth Millikan's *Language, Thought, and Other Biological Categories* (Cambridge, MA: The MIT Press, 1986) and David Papineau's *Philosophical Naturalism* (Oxford: Blackwell, 1993), which defend versions of the biological theory; Daniel Dennett's *Brainstorms* (Cambridge, MA: The MIT Press, 1981) and *The Intentional Stance* (Cambridge, MA: The MIT press, 1987), which contain essays developing his own unique approach to the problem; and Robert Cummins' *Meaning and Mental Representation* (Cambridge, MA: The MIT Press, 1989), which critically surveys the various naturalistic theories. Popper's critique of causal accounts of meaning is developed in "Language and the Body-Mind problem," in *Conjectures and Refutations: the Growth of Scientific Knowledge* (New York: Harper and Row, 1968) and at greater length in Karl Popper and John C. Eccles, *The Self and Its Brain* (London: Routledge and Kegan Paul, 1977).

Different versions of eliminative materialism are articulated and defended by Paul Churchland in "Eliminative Materialism and the Propositional Attitudes," available in the Chalmers and Rosenthal anthologies cited at the end of earlier chapters, and at length in his *Scientific Realism and the Plasticity of Mind* (Cambridge: Cambridge University Press, 1979), and by Stephen Stich in *From Folk Psychology to Cognitive Science: The Case Against Belief* (Cambridge, MA: The MIT Press, 1983). The objection that eliminativism is ultimately self-refuting is developed by Lynne Rudder Baker in *Saving Belief: A Critique of Physicalism* (Princeton, NJ: Princeton University press, 1987).

The argument about the indeterminacy of material representations was inspired by some remarks made by Heqbert McCabe in his "The Immortality of the Soul: The Traditional Argument," in Anthony Kenny, ed., *Aquinas: A Collection of Critical Essays* (Garden City NY: Anchor Books, 1969). Haldane develops his argument about concepts in several places, most accessibly in J. J. C. Smart and J. J. Haldane, *Atheism and Theism*, second edition (Oxford: Blackwell, 2003), which also contains a criticism of Dennett's strategy of homuncular decomposition along the lines suggested in the text.

That propositions cannot be identified with anything material is elegantly demonstrated by Alvin Plantinga in chapter 6 of his *Warrant and Proper Function* (New York: Oxford University Press, 1993). That

the abstract and universal character of the objects of thought concepts and propositions - entails that thinking cannot be a material operation is an idea as old as Plato and Aristotle. A popular contemporary version of the argument is to be found in Mortimer Adler's *Intellect: Mind Over Matter* (New York Collier Books, 1990). A more rigorous presentation is given by David Oderberg in "Hylomorphic Dualism," forthcoming in *Social Philosophy and Policy*, vol.22, no.2 (June 2005). James Ross's argument is developed at length in "Immaterial Aspects of Thought,' *The Journal of Philosophy* 89: 136-50 (1992). Kripke explains "quaddition" in *Wittgenstein on Rules and Private Language* (Cambridge, MA: Harvard University Press, 1982).

