
Minds, Brains, and Multiple Realizability

Questions about minds are as ancient as any in philosophy. But the mind–body problem as it exists today first appeared only in the middle of the twentieth century. The current mind–body problem does not concern the issue of whether minds are part of the physical world just as are aardvarks, hearts, rocks, cigarettes, heat, and the like. The problem is not *whether* minds are part of the natural world, but *how* they are.

Although the idea that minds are part of nature is not itself new, until quite recently there were no theories that could make sense of the contention. As a result, it was difficult even to imagine how mental phenomena could be natural phenomena.¹ Many factors contributed to the philosophical and scientific milieu that makes a naturalistic theory of mind credible. Advances in biology (e.g., discovery of the double-helix structure of DNA), psychology (e.g., replacement of behaviorism with cognitive psychology), mathematics (e.g., computability theory)—not to mention the advent of technologies suitable for manipulating the world as theorized by these new accounts, as well as a healthy dose of science fiction—all helped to set the stage for credible physicalist theories about minds. And, of course, the nascent neurosciences began to reveal that the complexity of the brain is of a scale not previously imagined.

At last we could begin to see how minds might be sophisticated physical phenomena. As Owen Flanagan writes, “we do understand how physicalism can be true. It can be true if every mental event is realized in the brain” (1992: 93). So the current mind–body problem

does not concern the general possibility of a naturalistic theory of mind. Instead, at issue is what the world—minds, brains, organisms, environments, and so forth—must be like if a naturalistic theory of mind is true:

the mind–body problem—*our* mind–body problem—has been that of finding a place for the mind in a world that is fundamentally physical. The shared project of the majority of those who have worked on the mind–body problem in the past few decades has been to find a way of accommodating the mental within a principled physicalist scheme, while at the same time preserving it as something distinctive—that is, without losing what we value, or find special, in our nature as creatures with minds. (Kim 1998: 2)

Theories are plentiful. They compete, for example, on the basis of which best complement other natural theories, and which give the most satisfactory or satisfying accounts of a range of mental phenomena. Of course there are still some theorists who doubt that mental phenomena are exhausted by physical phenomena. But they are not typically vexed by what a physicalist theory of mind would look like.² Instead, they simply contend that the world does not in fact work, perhaps could not work, in the way that physicalist theories of mind require. One might even believe that physicalism is correct up to a point, but that it leaves an unexplained remainder, as David Chalmers argues (1995, 1996a). So the important issues for a naturalistic philosophy of mind are how it requires the world to be, and whether the world lives up to those expectations.

1 Identity Theory and Putnam’s Intuition

The most rudimentary naturalistic theory of mind is the mind–brain identity theory: minds are identical to brains. In J. J. C. Smart’s words, “Sensations are nothing over and above brain processes” (1959: 163). Identity theory was the first serious contender for a mechanistic theory of mind.³

The slogan “sensations are brain processes” is admittedly a bit crude. For one, it is unclear whether identity theories should be thought of in terms of the identity of organs, or states, or processes, or events, or properties.⁴ But the core idea is so simple that it can hardly be refined: mental states, processes, events, or properties *just*

are brain states, processes, events, or properties.⁵ The merits of identity theory are equally plain. Smart advocated identity theory on the grounds that it is parsimonious and ontologically modest; it does not require us to posit new sorts of properties, events, states, and so forth.⁶ Thus it is thoroughly naturalistic; it appeals only to the ontology of the natural sciences. And identity theory is the only theory with a robust explanation of how mental phenomena can cause physical phenomena: “It makes mental causation entirely unmysterious: Mental causation turns out to be a species of physical causation” (Kim 1996: 56). If mental phenomena are identical to physical phenomena, then no distinctive question about mental causation ever arises. This is no small matter.

In spite of its merits, the identity theory was quickly displaced as the primary form of materialism. Its deposer is Hilary Putnam’s *functionalism*, the ontological thesis that mental states are not identical to physical states of brains but are instead *realized* or *instantiated* by physical states of brains. The main arguments against identity theory and for functionalism all originate from the single intuition that some creature could have sensations without having brain states like ours, or any brain at all. Putnam supposed—and many have followed him—that this intuition demonstrates the implausibility of identity theory:

Consider what the brain-state theorist has to do to make good his claims. He has to specify a physical-chemical state such that *any* organism (not just a mammal) is in pain if and only if (a) it possesses a brain of suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusc’s brain (octopuses are mollusca, and certainly feel pain), etc. At the same time, it must *not* be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extraterrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may *be* pain.

It is not altogether impossible that such a state will be found. . . . But this is certainly an ambitious hypothesis. (1967, in Putnam 1975c: 436)

Putnam takes the identity theorist to be committed to the view that every possible conscious creature must be capable of having states

just like our brain states. As an alternative, he proposes that conscious states can be realized by various biological or nonbiological states, such that creatures lacking brains and brain states like ours may nevertheless be conscious. At least, it seems probable that mental states can have *multiple realizations*.⁷ Functionalism, it is argued, is superior to identity theory because the latter cannot accommodate the overwhelming likelihood that mental states are multiply realizable.

Of course identity theory's sparse form does not leave much opportunity for nuance: mental states are identical to brain states. But as I suggested earlier, a theory of mind rests also on what it says about the world. I contend that the world is as identity theory takes it to be rather than as functionalism takes it to be. Although there is little room to amend identity theory itself, much can be said about how identity theory views the world. In this case, I need to explain how identity theories view multiple realizability.

My strategy is that of divide and conquer. I distinguish four interpretations of the multiple realizability intuition. I argue that the usual ways of construing multiple realizability are much stronger than can be supported by Putnam's intuition alone and should not be admitted. And the plausible forms of multiple realizability do not impugn the prospects for a mind-brain identity theory.

Arguments from multiple realizability purport to show that functionalism has the theoretical merit of being more general than identity theory. If two theories are comparable in other ways but one explains more than the other, the more widely applicable theory is to be preferred. Generality is a good thing. Functionalism, because it allows for multiple realizability, can explain the mentality of more kinds of things; specifically, it is not restricted to those creatures with brains like ours. But generality is a matter of degree. What degree of generality is appropriate to a theory of mind? That is, to what degree are mental states multiply realizable? There is no getting around the fact that identity theory cannot handle the wildly different realizations of mental states that functionalism so easily accommodates. In the end the identity theorist will have to settle for identities that are in some way restricted. The questions we must ask, then, are whether restricted identities are bad or compromise identity theory and

whether the unrestricted multiple realizability licensed by functionalism is desirable. We need to know how much generality is required if we are to assess whether or not identity theory is compatible with multiple realizability. The answer, to pervert Daniel Dennett's (1984) expression, is that identity theory can accommodate the varieties of multiple realizability worth wanting.

2 Varieties of Multiple Realizability Worth Wanting

Recently there has been renewed attention—largely critical—to multiple realizability. Some have aimed to defuse the threat that multiple realizability seems to present for various theses of explanatory “reduction” (Bickle 1998; Sober 1999; Shapiro 2000, forthcoming). Others have been concerned to resolve the difficulties that multiple realizability is alleged to introduce for explanations of mental causation (Kim 1998; Heil 1999). Still others have questioned the empirical evidence for multiple realizability, and whether it constitutes an obstacle to the practices of neuroscientists (Bechtel and Mundale 1999; Bechtel and McCauley 1999).⁸ For the most part the authors mentioned treat multiple realizability as a structural constraint on explanations of mind. Each attempts to show that within the constraints of multiple realizability we can nevertheless justify “reductive” explanation, account for mental causation, carry on neuroscientific practice, and so forth. Some, but not all, believe that their conclusions help to support some version of identity theory.

Certainly structural features are an important aspect of arguments based on multiple realizability; but what makes multiple realizability such a thorn is that it enjoys a presumptive advantage that is based on the intuitive appeal of its content. That content, in turn, dictates the structural constraints. If we could find reason to question the content of the multiple realizability intuition, then perhaps we would not have to theorize within its structural constraints. And I think we have such reasons.

Putnam's intuition is that we should not expect every kind of conscious creature to have a brain with states just like our own. I find this intuition plausible. It seems to accord with the information that we get from biologists and cognitive ethologists. There is at least

good reason to think that human beings are not the only conscious creatures, even on our little planet. But it is a far stretch from Putnam's intuition to multiple realizability as it is usually entertained. Consider these samples from the spectrum of multiple realizability:

Weak MR. At least some creatures that are not exactly like us in their physical composition can be conscious.

SETI MR. Some creatures that are significantly different from us in their physical composition can be conscious.⁹

Standard MR. Systems of indefinitely (perhaps infinitely) many physical compositions can be conscious.

Radical MR. Any (every) suitably organized system, regardless of its physical composition, can be conscious.

The forms of multiple realizability most commonly encountered in the literature are what I have here dubbed standard MR and radical MR. Examples are numerous:

[T]he mere fact that a creature's physical states are radically different from ours does not in itself preclude them from being realizations of mental states. (Shoemaker 1981b, in 1984: 280)

Creatures on this planet to which we unhesitatingly ascribe a range of mental characteristics differ *biologically* from us in endless respects. (Heil 1992: 133)

[T]here are infinitely (indefinitely) many possible physical "realizations" of pain: C-fibers, silicon fibers, et cetera. (Bealer 1994: 187)

Mental property *M*—say, being in pain—may be such that in humans C-fiber activation realizes it but in other species (think of octopuses and reptiles) the physiological mechanism that realizes pain may well be vastly different. And perhaps there are biological systems—at least no laws of nature rule out such a possibility—that are not carbon- or protein-based, and there can be electromechanical systems, like robots in science fiction, that are capable of having beliefs, desires, and even sensations. All this seems to point to an interesting feature of mental concepts: They include no constraint on the actual physical/biological mechanisms or structures that, in a given system, realize or implement them. (Kim 1996: 74–75)

For more than three decades antireductionists have argued that radically different physical systems could realize identical mental kinds. . . . They often defend the premise of multiple realizability with thought experiments

involving silicon-based extraterrestrials, computers, androids, robots, and other brainless science fictional beings. (Bickle 1998: 114)

To be sure, if functionalism is correct then standard MR is likely, and perhaps even radical MR. But if a variety of multiple realizability is to be the basis for an argument against identity theory, then it will have to be a form of multiple realizability whose plausibility does itself not depend on the truth of functionalism. This is plain enough, for if the plausibility of multiple realizability depends on the plausibility of functionalism, then the argument from multiple realizability only repeats the assertion that functionalism is plausible.

Let us ask: Does anyone have the pretheoretic intuition that mental states admit standard MR—that systems of indefinitely (perhaps infinitely) many compositions can be conscious? Of course, any intuitions are colored by some amount of theory. But functionalism is the particular theory about which we are now concerned; our intuitions regarding what is plausible have not been immune to the proliferation of functionalist theories. I contend that standard MR and functionalism go together. Standard MR is not part of any argument against identity theory that is independent of substantial metaphysical claims. Of course, there is nothing wrong with arguing against identity theory from the combination of functionalism and standard MR; but that would not be an argument that could generate a presumptive judgment against the plausibility of identity theory, as the argument from multiple realizability has been thought to. If the multiple realizability argument is to be a burden to identity theory then it must itself be plausible independent of functionalism.

I am not going to rest my conclusion only on this “he-said-she-said” argument. But notice that the case is not as flimsy as it at first seems. It is extremely difficult to see how someone who endorses standard MR can resist the thesis of radical MR, that any suitably organized system can be conscious. As Fred Adams puts the point, “If the state is realizable by indefinitely many [physical-chemical] systems, then why not all? The natural reply is that some things simply are not put together in the appropriate way to be in pain. Hence, we doubt that trees and rocks and grass, etc., feel pain (otherwise we might stop mowing our lawns)” (1979: 159). If standard MR were itself

plausible, its supporters seem to be committed to the stronger thesis of radical MR. I am not claiming that standard MR entails radical MR. It does not. But unless some explanation can be given that justifies constraints on realizers of mental states, the considerations that justify belief in standard MR will also tend to support radical MR. What Adams calls “the natural reply” to block the slip from standard MR to radical MR is not available because to explain why “some things simply are not put together in the appropriate way” requires invoking a substantial theory. But the multiple realizability arguments, if they are to be a problem for the identity theory, are supposed operate prior to adopting any such theory.

We can put this problem in the form of a dilemma: Either standard MR is a consequence of prior metaphysical commitments, or it is not. If standard MR is the consequence of a prior metaphysical commitment, then it is question-begging with respect to the identity theory. So it cannot be a consequence of prior metaphysical commitment. But if standard MR is not a consequence of prior metaphysical commitment, then it looks as though it is a premise in a conceivability argument. Yet if it is a premise in a conceivability argument, then there seems to be no reason that radical MR should not also be conceivable. This will be so even if the evidence for the conceivability claim is that some things that are physiologically different from us (e.g., octopi) in fact have sensations (e.g., pain). That evidence, if it is evidence for anything at all, is also evidence for radical MR. It seems that the grounds for believing in standard MR also justify belief in radical MR. But belief in radical MR is more than dedication to theoretical generality. Radical MR is a substantial metaphysical thesis that is not plausible independently of the rejection of identity theory. Radical MR cannot be a premise in an argument against identity theory.

If standard MR is not question-begging, then some reason must be given for drawing a line between standard MR and radical MR. But it is not clear that any such explanation could be made without invoking a theory of the nature of mind.¹⁰ And if one must invoke a philosophy of mind to justify limiting realizers, then to whatever extent the justification is incompatible with identity theory it is question-begging. A theory-based multiple realizability argument will

not have the presumptive force that multiple realizability arguments assume.

Needless to say, I dismiss as hogwash anyone's claim to have metaphysically neutral and prefunctionalist intuitions in favor of radical MR. This is brazen, I know. To see why I can make this claim, it is important to keep in mind how the multiple realizability argument is supposed to work. Multiple realizability is a claim about the generality of accounts of the ontology of minds. Some versions of functionalism, if true, would admit radical MR. But that is not the question at hand. Rather, we want to know whether it is a desiderata on theories of mind that they permit radical MR. Is that degree of multiple realizability a virtue that we expect any such theories to have? Like you, I seem to be able to imagine that things radically different from human beings, with silicon chips rather than neurons, say, could turn out to have conscious mental states. What I am imagining is that some particular substantial account of mental ontology is true, and that on that theory things with silicon chips rather than neurons have minds. Of course I don't have to articulate the theory; most people most of the time don't think about that theory when they imagine or enjoy the adventures of affectionate robots or surly talking hats.¹¹ And even so I'm not at all confident that we're imagining that radical MR is true—that *anything* could be conscious—rather than standard MR. Radical MR is the provenance of the kind of fantasy that invokes magic to animate trees and teacups, rather than the futuristic robots and aliens whose familiarity in modern storytelling seems to underlie many multiple realizability intuitions. Maybe robots can be conscious. But it is not a prior constraint on any account of conscious experience that it allow radical MR. Yet this is what radical MR asserts when it is used in a multiple realizability argument against identity theory: It maintains that we have prior commitment to ensuring that any theory of mind will be general enough to account for the minds that robots and automata could have. But that cannot be right. If robots can have minds, that is a substantial discovery; it is not something guaranteed merely by a commitment to theoretical generality.¹²

Some readers will fail to be moved by the powerful argument against standard MR and radical MR just offered, the "Argument

from Nuh-uh.” Let us try another approach, for we identity theorists have many tools at our disposal. The strategy, recall, is divide and conquer—or, more accurately, divide and unify. So far I have been making the case that the stronger forms of multiple realizability should not enjoy our presumptive support. But suppose that, for one reason or another, we are obliged to show that identity theory is compatible with standard MR and radical MR. What resources can an identity theorist offer? We may reply, with Jaegwon Kim:

[T]he fact that two brains are physico-chemically different does not entail that the two brains cannot be in the “same physico-chemical state.” Even if we disallow the *ad hoc* creation of new states by forming arbitrary disjunctions, the remaining possibilities are indeed limitless. . . . If the human brain and the reptilian brain can be in the same “temperature state,” why can they not be in the same “brain state,” where this state is characterized in physico-chemical terms? . . . the mere fact that the physical bases of two nervous systems are different in material composition or physical organization with respect to a certain scheme of classification does not entail that they cannot be in the same physical state with respect to a different scheme. (1972, in Block 1980a: 234–235)

Multiple realizability is typically couched in terms of what sorts of systems are capable of having mental states, and I have followed suit. But the fact—if it is a fact—that many different systems can have the same kinds of mental states does not show that they do not all do so in virtue of having something in common. An easy way to think of this scenario is in terms of systems sharing some properties but not others. Two systems may differ with respect to the properties by which neurological states are categorized, but may nevertheless share other properties that can ground their classification as the same kind of mental state. As Frederick Adams notes, “Just because two [physical-chemical] systems are different kinds of stuff does not mean that they do not share some identical property-kinds” (1979: 158; see also 1985). The identity theorist is free to maintain that creatures that are made of different stuff than human beings, and thereby have different physical states than we do, nevertheless share some properties that we have. Specifically, the usual application of multiple realizability arguments leaves open that these diverse systems may be like us with respect to the properties (whatever they

turn out to be) on which conscious experience depends. Multiple realizability intuitions at most suggest that conscious mental states can be realized in various systems, but they do not support the contention that such states are realized by different properties in different systems rather than by properties that the systems share. Call this the Kim–Adams reply.

Identity theory contends that mental states are identical to physical states; but that does not require that all mental states of a kind are identical to each other. The relationship between mental states of a kind is merely that they are the same with respect to those properties characteristic of the kind. And it is quite reasonable to think that at least some creatures that are different from us—with different brains, say—could nevertheless have states that have the relevant properties in common with us. If this is right, then weak MR is no threat to identity theory, for it is no problem if some creatures not physically identical to us have conscious states as long as they share some properties with us.¹³ Lawrence Shapiro makes a similar point when he argues that realizers count as genuinely multiple only if they differ in causally relevant ways, “in properties that make a difference in how they contribute to the capacity under investigation” (2000: 644).

Recall again the divide and conquer strategy being pursued. We have seen that a range of claims figures in multiple realizability arguments. The stronger claims, I first argued, are not plausible independently of a commitment to functionalism. Next I argued that the Kim–Adams reply may be employed to show that identity theory can accommodate a broader range of cases than is usually supposed—at least those of weak MR. This reply does not guarantee that all creatures capable of consciousness in fact have some property or properties in common. It demonstrates only that Putnam’s intuition does not by itself provide sufficient reason for abandoning identity theory.

Applied to weak MR, the Kim–Adams reply shrugs off multiple realizability concerns. On the other hand, by offering the Kim–Adams reply to standard MR or radical MR, the identity theorist digs in: either the thing shares some properties in common or else it does not have mental states after all. But the Kim–Adams reply has not

persuaded many standard MR loyalists, and we might hope that more can be accomplished than merely reporting a clash of intuitions. Rather than insist on the point, let's see what more the identity theorist can say. The Kim–Adams reply is one part of the total defense of identity theory from multiple realizability arguments, but not the only part.

3 Empathy and Other Species

Consider the form of multiple realizability that I am naming after the Search for Extraterrestrial Intelligence project, SETI MR. Let us imagine that some creature that is quite different from us physiologically, even neurophysiologically, could nevertheless be conscious. How different could such a creature be? Suppose for a moment that the Kim–Adams reply only gets us so far; beyond that point the creatures we imagine are so different from human beings that it is hard to see how their mental lives could be based in properties they share in common with us.

Notice how much is being supposed. These creatures would have to be very different indeed if they are such that we can rule out—ahead of time—the possibility that their “brains” could have any properties of the same kinds that our brains have. This is much more than Putnam's intuition alone suggests; but it does seem to be the idea behind the legion of Martians and other aliens that have populated philosophical discourse for some time now. For the sake of argument let us say that there could be such creatures; if so, they are squarely in the range of SETI MR. Explaining how identity theory accommodates SETI MR requires us to look more carefully at what is being claimed by versions of multiple realizability.

We should distinguish two ways of interpreting any multiple realizability claim. As formulated above, SETI MR is the thesis that some creatures that are significantly different from us in their composition could be capable of having some conscious states or other. A stronger claim would be that some creatures that are significantly different from us in their composition could be capable of having *exactly the same kinds* of conscious states that we have; call this the *empathetic* interpretation of multiple realizability. According to empathetic

versions of multiple realizability, other creatures could have mental states that are of exactly the same kinds as our own.

The term “empathetic” is perhaps less than perspicuous. I have in mind that if a creature has, for example, pains that are just like our own, we might be able to imaginatively identify with its experiences. If a creature has states that are not like our own, we might not have a basis for imaginative identification. Of course I might in some sense sympathize or empathize with something like a tree, a building, or a nation, something that does not genuinely have conscious states in the sense at stake herein. I do not have a theory of empathy or sympathy to offer. My metaphor is loose, and I hope not too misleading. The point should be clear enough: the multiple realizability arguments require that other creatures can have exactly the same mental states that we do. Without empathetic multiple realizability, there is no argument against identity theory. For multiple realizability to be a problem for identity theory it is not sufficient that some wildly different creature have some conscious state or other; it must be that different creatures can have exactly the same—empathetic—kinds of mental states.

So we should take it that empathetic interpretations of multiple realizability are what most philosophers have in mind. This is strange because it runs counter to ordinary intuitions about sensations. Despite what Bill Clinton may say, none of us believes that he truly felt our pain. We expect that there are differences between individuals, even within individuals over time, that make it unlikely that one sensation is ever exactly similar to another. Of course, we generally allow that other people have some sensations of the same kinds that we do. So empathetic multiplicity is not entirely ruled out among human beings. But we expect limits. We do not, in general, believe that all human beings have all, only, and exactly the same kinds of sensations as each other. We say, for example, that artists and musicians—good ones anyhow—see and hear things that the rest of us do not. Late-night television demonstrates the widespread belief that some people (but only a few people) have an extra sense through which they perceive information about the future; presumably those senses are supposed to be accompanied by distinctive “seeing the future” sensations that I do not have.¹⁴ More mundane examples

are humans with reduced capacity for kinds of sensations, for example, persons with cortically based color deficiency (Hurvich 1981).

When we consider nonhuman animals, we become even more reticent in our empathetic attributions. Many of us believe that at least some nonhuman animals are conscious. We are pretty sure, for example, that domestic dogs have sensations. But are we committed to the belief that dogs have pains and itches of the same kinds as our own? Perhaps; perhaps not. We needn't stray far from the human animal before it seems to us that, say, although dolphins experience sensations, dolphin sensations are not exactly the same kinds we have. This, famously, is why Thomas Nagel chose bats as his example in his "What Is It Like to Be a Bat?" Bats are similar to us physiologically, but they seem to have quite disparate ways of being in the world: "Bats, although more closely related to us than those other species, nevertheless present a range of activity and a sensory apparatus so different from ours that the problem I want to pose is exceptionally vivid . . ." (1974: 438). Nagel goes on to explain that bats provide a terrestrial example of a "fundamentally *alien* form of life" and experience. My selection of dolphins, another echolocating mammal, is similarly motivated.¹⁵

When we consider creatures with physiologies very different from ours, our eagerness to attribute empathetic conscious states quickly dwindles. It may have seemed to David Lewis that Martians could have sensations like ours. But it does not seem to me that even avian or piscine sensations are quite like ours, although I suppose I am inclined to believe that birds and fish have sensations of some kinds. This might be wrong. But if it is we will have to do much work to discover that other creatures have sensations empathetically like our own. It is certainly not a platitude of common-sense psychological reasoning that all conscious terrestrial animals have experiences that are exactly, empathetically, of the same kinds. And if they do not, then multiple realizability never gets a foothold. For the claim that pain is multiply realized cannot be any more plausible than the claim that pain is had by various creatures to begin with. "According to functionalists, we have vivid intuitions to the effect that mental states are distributed across an extremely broad spectrum of biological

species,” Christopher Hill (1991: 45) writes. Probably many creatures have some conscious mental states or other, but I doubt that empathetically the same mental state kinds are common. I contend that we do not attribute empathetically the same mental state kinds to very many creatures—not even to some terrestrial creatures that probably share many properties with us. Standard MR and radical MR are not at all plausible in their empathetic forms.

All this is to say that from the position of pretheoretic intuition—remember, that is the position from which generality arguments from multiple realizability claim to operate—it appears that *species-specific* variation of mental states is to be expected, rather than multiple realizability of empathetically the same states. We expect only very similar creatures to have very similar sensations. To the extent that a creature differs from us, so will its experiences. Pain in human beings may not be exactly similar to pain in dogs, dolphins, or Martians; pigeons and turtles with extra photoreceptors might well have color sensations that we do not.¹⁶ Species-specific identities are compatible with our intuitions about what things have minds and compatible with the identity theory.¹⁷

The move toward species-specific identities is not new. Paul Churchland has advocated this tactic (P. M. Churchland 1979, 1982), as has Berent Enç (1986).¹⁸ And from the start David Lewis (1969) and Jaegwon Kim (1972) each defended identity theories from multiple realizability arguments on the grounds that they need to posit only local or species-specific identities:

Let us assume that the brain correlate of pain is species-dependent, so that we have generalizations like “Humans are in pain just in case they are in brain state *A*,” “Canines are in pain just in case they are in brain state *B*,” and so on. These species-dependent correlations . . . clearly do warrant—at least they are not inconsistent with—the identification of *human pains* with *human brain state A*, *canine pains* with *canine brain state B*, and so on. . . . There seems no reason to suppose that species-specific identities are insufficient for the identity theory. (Kim 1972, in Block 1980a: 235)

Putnam, recall, takes the identity theory to claim that every creature capable of a certain kind of mental state must also be capable of exactly the same kind of physical brain state, and he asserts that many creatures do in fact have the same kinds of mental states. Lewis

(1969) argues that a more reasonable view is that mental states are identical to distinct states in various kinds of creatures. Likewise, although Terence Horgan does not himself advocate the view, he writes that “[t]he appropriate form of the identity theory is one that also accommodates species-relative multiple realization—despite the fact that multiple-realization considerations are commonly thought to undermine type-type psychophysical identities altogether” (1997: 165).

Of course a good deal more needs to be said about how species-specific identities are supposed to work, and we’ll return to that topic shortly. But even if local or species-specific identities can be explained, restricting identities to particular species is not a cure-all for identity theory. Differences between members of a species present a case of multiple realizability that has troubled some philosophers. Horgan paints a particularly apocalyptic picture according to which neurobiological anarchy underlies well-behaved psychological kinds; there is indeed trouble for identity theory if mental states are “radically multiply realizable, at the neurobiological level of description, even in humans; indeed, even in individual human beings; indeed, even in individual humans . . . at a single moment” (1993: 308). And John Bickle notes, “Surely this much domain specificity is inconsistent with the assumed generality of science” (1998: 124). Individual differences and intraindividual differences do not tolerate appeals to species-specificity; we cannot have species of one. Even William Bechtel and Jennifer Mundale, who do not consider intraspecies differences to be a difficulty for identity theory, note that the fan of multiple realizability will find variation to be ubiquitous: “[I]t is well known that even within a species brains differ. Even within an individual over time there are differences (neurons die, connections are lost, etc.). Thus multiple realizability seems to arise within species (including our own) and even within individuals” (Bechtel and Mundale 1999: 177).

But identity theorists do not need to invoke yet more localization within species and individuals; we have another tool. Species-specificity is one part of the divide and conquer strategy. Some mental states are not multiply realized because they are species-specific. But we can also unify. Some states are not multiply realized

because all their purportedly various instances in fact share properties in common. That is, we may return to the Kim–Adams reply—it is quite effective for cases of within-species and within-individual differences, for it is quite likely that members of species share properties with one another.

Those who are concerned about individual and intraindividual differences seem to have in mind two sorts of examples. On the one hand, there is a myriad of near-miraculous case studies wherein persons with massive neuropathologies nevertheless live normal or near-normal lives. These include the recovery of stroke victims, as well as more amazing cases of individuals who have only a small amount of viable cortex. On the other hand, there is the ordinary variance in neural activity over time, such as the well-known changes to motor cortex depending on recent use.¹⁹ Many of us are familiar with diagrams in psychology textbooks that illustrate changes in primate cortex mapped to the fingers caused by exercise or amputation of the fingers (figure 1.1). Of course these are only a tiny fraction of cases. Most subjects who suffer massive neurological traumas do not have uninhibited behavior, and much neural function appears to be reasonably localized and stable over time. Surprising case studies are, well—surprising. But they do not show that exactly the same mental state is multiply realized, nor that similar mental states have wildly different realizations. Rather than supporting multiple realizability, these cases suggest that we do not understand very well how the brain works—how to individuate brain processes, events, states, and properties.

The considerations we have been entertaining so far leave plenty of room for identity theory to accommodate Putnam’s intuition. If functionalism is the correct theory of mental states, radical MR and standard MR are not unreasonable; but they are implausible, independent of a commitment to functionalism. So even if it is true that identity theory does not permit that degree of variability in realizers of mental states, it is no defect of the theory. Although I have contended that few creatures have empathetically the same mental states that we have, if they do then the Kim–Adams reply reminds us that even biologically diverse realizing systems can share some properties with one another. Mental state kinds are probably as widespread as

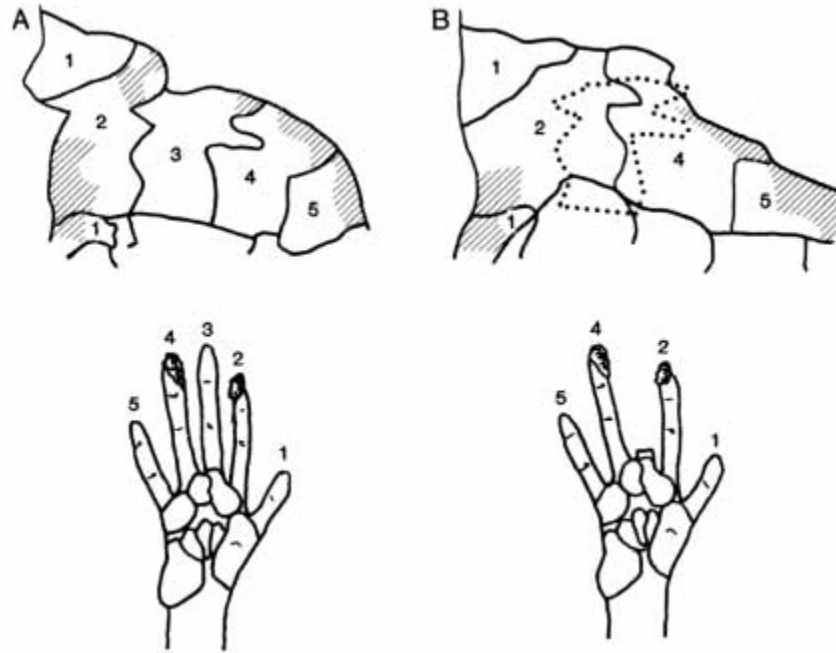


Figure 1.1

Cortical plasticity. A portion of the cortex of an owl monkey is organized as a somatosensory “map” of the hand in which each finger is represented by a distinct area of tissue (A). Within weeks following the amputation of the middle finger the map reorganizes (B) and the area that previously represented the amputated digit is now “fully partitioned by areas representing the two adjacent digits” (Fox 1984: 821). Reprinted with permission from J. Fox, “The brain’s dynamic way of keeping in touch,” *Science* 225 (1984): 820–821. Copyright 1984 American Association for the Advancement of Science.

weak MR and moderate forms of SETI MR contend. But that is no problem for identity theory, because the Kim–Adams reply explains how identity claims can cover more creatures than is typically supposed.

In those cases where identity theory declines to attribute a kind of sensation to members of a species, it does so because the creatures in question are sufficiently different to make it improbable that they share relevant properties with us. But those are just the cases for which our pretheoretic intuitions tell us that it is unlikely they share

exactly similar conscious state kinds with us. Species-specific identities are what we would expect if mental states are widespread in the ways that weak MR and SETI MR contend: Alien beings may have conscious states of a sort; and those are identical to some of their physical states, just as our conscious states are identical to some of our physical states. Identity theory thus accommodates multiple realizability claims to just the extent we should want.

And this is just what we should have expected all along. Consider everybody's favorite example, a carburetor. Let us remind ourselves a thing or two about carburetors (see figure 1.2). Carburetors mix air and fuel, and control the air-to-fuel ratio, in combustion engines such as those found in many commercial automobiles. Liquid gasoline must be vaporized before it is introduced into the combustion chamber of an engine. Fuel is drawn into the carburetor where a partial vacuum and the temperature of the intake manifold cause the fuel to vaporize. The continuous partial vacuum, that is, the lowered pressure, in the intake manifold while the engine is running is the key to the operation of a carburetor.

One cause of the vacuum is the withdrawal of the piston caused by combustion. In addition: "To make sure that there is always enough vacuum at the carburetor to draw fuel out of the float bowl, the carburetor throat narrows down and then opens up again to form a shape known as *venturi*. As air passes through a venturi, air speed increases and pressure drops. A vent at the top of the float bowl allows air at atmospheric pressure to push fuel through a passage and into the low pressure area of the carburetor throat at the venturi" (*Reader's Digest* 1981). To the best of my knowledge this is more detail about how carburetors work than has ever been explained in a philosophical argument. (So I will skip the nuances of throttle plates, by which the vacuum in the venturi is made to regulate the air-to-fuel ratio.)

Are carburetors examples of multiply realized kinds? Of course. There are many manufacturers of carburetors, which are made of various materials and of various designs. Some have more than one venturi, for example. Carburetors allow weak MR. Depending on what you think about the differences between carburetors that have single or multiple venturi, perhaps they allow SETI MR. But there is

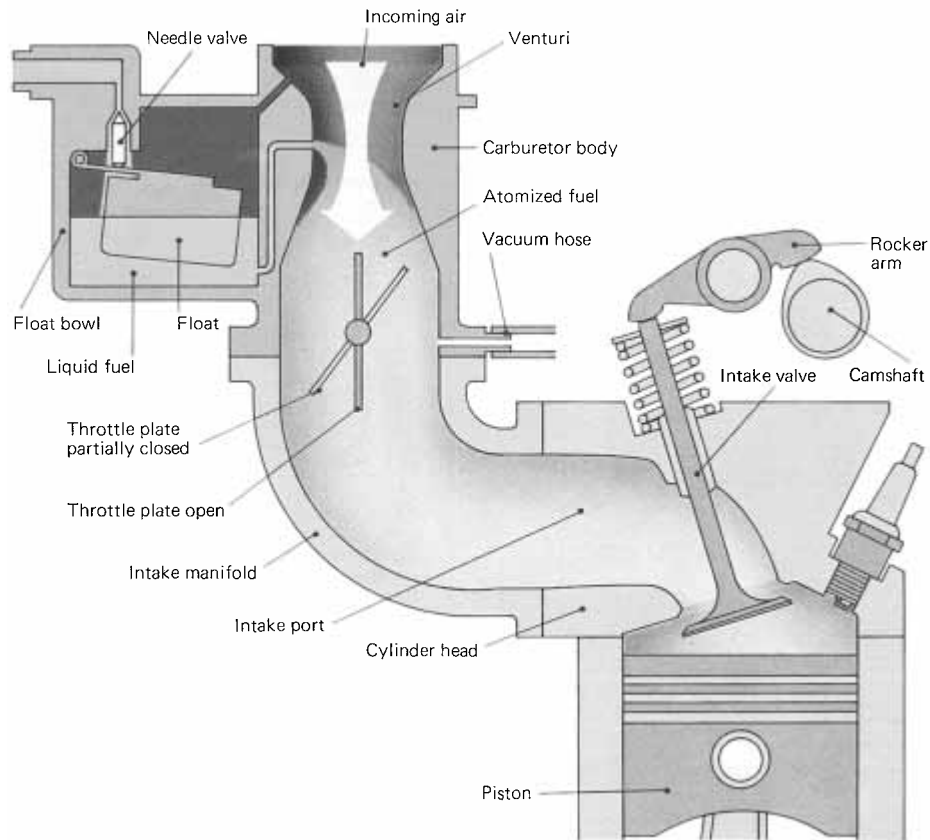


Figure 1.2

A simple carburetor. Reprinted with permission from *Complete Car Care Manual*, © 1981 The Reader's Digest Association, Inc., Pleasantville, N.Y., www.rd.com. Illustration by Dale Gustafson.

a single structural feature that all carburetors have: they all have venturi. If some device for mixing air and fuel does not work by the venturi principle, if it does not have venturi, then it is not a carburetor. In particular, fuel injectors are not fancy carburetors; they do not have venturi.²⁰

4 Granularity and the Evidence for Multiple Realizability

Both the Kim–Adams reply and Kim’s and Lewis’s defenses of species-specificity have been part of identity theorists’ arsenal for some time. Yet the idea that multiple realizability is a problem for identity theory persists. It would be nice if we had an account that explains the widespread dedication to forms of multiple realizability, while demonstrating that they are no threat to identity theory. Recently William Bechtel, Jennifer Mundale, and Robert McCauley have given just such an account. They suggest that multiple realizability arguments against identity theory depend on misunderstanding the extant neurophysiological evidence. Bechtel and Mundale (1999) are particularly concerned to show that the facts of multiple realizability do not undercut the potential contributions of neuroscience to the understanding of cognition, while Bechtel and McCauley (1999) employ similar considerations to defend a version of identity theory itself.

The multiple realizability that Bechtel and his collaborators have in mind is standard MR. In a telling passage, Bechtel and Mundale write:

The claim of multiple realizability is the claim that the same psychological state can be realized in different brain states. . . . Taking this a step further, many philosophers became convinced that the same mental activities could be realized in the brains of aliens with radically different composition from ours. The upshot of these speculations about artificial and alien minds is a metaphysical claim that mental processes are operations themselves, and are not identified with whatever biological or other substances realize them. (1999: 176)

Although this passage begins by stating a version of weak MR, it immediately admits that systems of radically different composition may be taken to have minds—which is SETI MR. Then, by attributing to multiple realizability the consequence that minds can be identified independently of “whatever biological or other substances realize them,” Bechtel and Mundale indicate that they are concerned with full-blooded standard MR.

A common element in the arguments advanced by Bechtel and his collaborators is that the claims of multiple realizability depend on an equivocation of the “grain” of mental states and brain states. The

grain of a mental state or brain state is the degree of generality with which it is identified. The taste of chocolate is presumably a finer-grained sensation than the taste of sweetness, for example.²¹

Bechtel and Mundale diagnose that multiple realizability arguments against identity theory depend on an equivocation of granularity. If we individuate mental state kinds coarsely it might seem that biologically diverse creatures are capable of sharing psychological states. Putnam seems to have in mind relatively coarse-grained states, for he offers only that psychological laws “of a sufficiently general kind” will be species independent (1967, in Putnam 1975c: 437). But, according to Bechtel and Mundale, in continuing on to claim that the creatures under consideration are biologically diverse, Putnam is illicitly contrasting coarse-grained psychological states with fine-grained brain states:

A human’s psychological state and that of an octopus might well be counted as the same insofar as they are associated with some general feature (such as food-seeking behavior, in the case of hunger). But with respect to other considerations, a human psychological state may be considered different from that of an octopus, even if we limit the scope to mere behavior. . . . the assertion that what we broadly call “hunger” is the same psychological state when instanced in humans and octopi has apparently been widely and easily accepted without specifying the context for judging sameness. (Bechtel and Mundale 1999: 203)

If this is right, the appeal of Putnam’s intuition is revealed to be an artifact of our failure to fix our mind–brain comparisons at a common granularity. It may seem that mental state kinds cannot be identical to brain state kinds if we are not considering states of matching granularity.

What is the appropriate grain size for sensations and brain processes? The suggestion is that there is no one answer; sensations and brain processes may each be considered with finer or coarser grain. Functionalists and identity theorists can agree that mental states occur at multiple granularities, some at one grain and others at other grains. The problem of determining the appropriate granularity for the brain state associated with a given sensation is part of what George Graham and Terence Horgan call the “grain project”: “Here is a large-scale, long-term, but potentially empirically tractable

project for cognitive science: to identify the specific causal role or roles associated with phenomenal states—to identify what, in some sense, those states do. . . . Let us call this scientific program the grain project, since it involves investigating the causal roles associated with phenomenal consciousness at several levels of detail or resolution” (2002: 65).²² Graham and Horgan suggest that a good theory of consciousness (“phenomenal states”) will integrate explanations across the various grains or levels. This is just what William Lycan (1987) argues for functionalism; and there is no reason to suppose that identity theory cannot follow suit. But whereas the functionalist will theorize at multiple functional grains, the identity theorist will operate at multiple biological grains.

At this point we are in a position to address the common complaint that, although the Kim–Adams reply and species-specificity could salvage identity theory philosophically, it is nevertheless empirically false. The idea seems to be that even if there are no current cases of multiple realizability, we expect neuroscientists to discover some any minute now. Bechtel and Mundale attribute the stubborn persistence of this belief, also, to confusions about granularity—in effect maintaining as Kim and Adams do that mental state realizers do share some properties. Bechtel and his collaborators point out that the advocates of multiple realizability seem to have ignored the extent to which neuroscientific research already depends on identifying morphologically similar and homologous brain structures across individuals and species. Simply put, the methods of present-day neurosciences depend on our ability to identify common brain structures across reasonably various species of creatures: “Historically . . . neuroscientific practice routinely involved identifying brain areas and processes across a broad range of species as belonging to the same type. These practices continue now. . . . Oddly, when they consider theories of mind–brain relations, philosophers seem to lose sight of the fact that the overwhelming majority of studies have been done on non-human brains” (Bechtel and McCauley 1999: 3–4). When the “same” mental state seems to be identifiable with diverse brain states, neuroscientists typically conclude that they are actually dealing with more than one mental state; they then proceed to look for psychological differences between the mental states.

Bechtel and colleagues survey an array of classic and contemporary studies using comparative neurobiology to identify brain areas by their anatomy or function, concluding that neuroscientific practice and data support mind–brain identity claims, sometimes across species. For example, they observe that Brodmann’s (1909) groundbreaking brain mapping is both cytoarchitectonic and comparative (figure 1.3). Brodmann used cellular staining techniques to distinguish brain areas according to the type and density of neurons. And Brodmann applied his technique not only to humans but to other species as well: “[I]n defending the claim of six different layers in the cortex, he reports preparations made from 55 species ranging over 11 different order of mammals. . . . When he turns to mapping the cortex, Brodmann presents maps for several other species in addition to humans: two lower monkeys (guenon and marmoset), lemur, flying fox, kinkajou, rabbit, ground squirrel, and hedgehog, using the same numbering systems to identify homologous areas in the different species” (Bechtel and Mundale 1999: 180). Bechtel his collaborators’ examples are not decisive, of course, but they suggest that that current and historical neurobiological data and practice do not foreclose the possibility of an identity theory.²³ Putnam was too quick to assert that there is neuroscientific support for multiple realizability. Current evidence does not reveal multiple realizability, and there is some reason to think the enterprises of neuroscientific investigation are premised on the hypothesis of brain-state identities.

This is exciting empirical work. Does the evidence go so far as to support identity theory? Bechtel and Mundale defer the question: “For the most part, we will have nothing to say about these speculative arguments [multiple realizability arguments against mind–brain identity], nor are we primarily concerned with the metaphysical claim” (1999: 176).²⁴ Although I am concerned with the metaphysical claim, I share Bechtel and Mundale’s hesitation to think that available evidence supports identity theory over functionalism. I doubt that the evidence is even suggestive in this respect, because I doubt that evidence by itself can settle metaphysical questions such as that between functionalism and identity theory. What Bechtel and his collaborators provide is nevertheless quite useful. They give us a

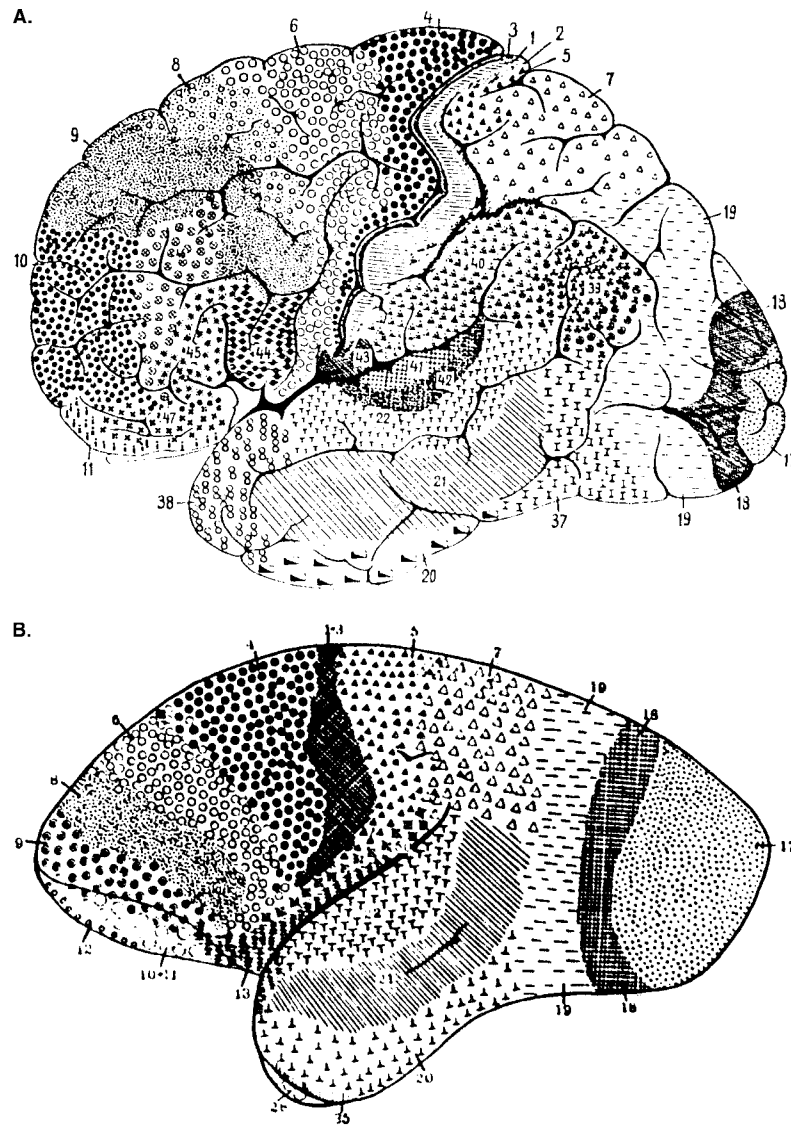


Figure 1.3
 Brodmann's brain "maps." (Brodmann 1909. Used by permission.) (A) Map of human cortex (lateral view). (B) Map of marmoset cortex (lateral view). As Bechtel and Mundale note, "Brodmann adapted the same numbering system as he used on the human brain for each species that he studied. In some species areas identified in the human brain were not identifiable, but each region of the cortex that was identified in a given species was associated with a region of the human cortex" (1999: 181).

way of understanding the lingering appeal of multiple realizability claims, even those as strong as standard MR; and they also give us the tools to deflect those concerns.

5 Species-Specific Identities and the Generality of Psychology

We now to return to the varieties of multiple realizability that I claim are compatible with identity theory. We need some assurance that the identities that we can get from such a theory are worthwhile.

The Kim–Adams reply to multiple realizability maintains that different mental states share properties after all. Mental states are multiply realizable, but mental properties are general. At the end of the day, this may prove true and thus be the identity theorists’ most powerful defense. However, I have taken a more modest stance, conceding for the moment that the Kim–Adams reply may have its limits; and I have argued that they are just the limits we expect. When a creature is so different that it cannot have states with the properties that our neural states have, then the Kim–Adams reply gives way and our identity claims are localized to each species. How much difference must be involved is not clear at this time. Perhaps, as Putnam thought, octopi are different enough.

How, then, does species-specific identity work? Lewis advocates the view that the same kind of mental state can be realized by various kinds of physical states in different creatures: “[Putnam] imagines the brain-state theorist to claim that all organisms in pain—be they mollusks, Martians, machines, or what have you—are in some single common nondisjunctive physical-chemical brain state. Given the diversity of organisms, this claim is incredible. But the brain-state theorist who makes it is a straw man. A reasonable brain-state theorist would anticipate that pain might well be one brain state in the case of men, and some other brain state in the case of mollusks” (1969, in Block 1980a: 233). Lewis thus holds that psychological kinds occur cross-species but that realizers of psychological kinds are species-specific. This is the disjunctive view; it identifies each mental kind with a disjunction of physical kinds (figure 1.4). David Lewis and David Armstrong have defended the disjunctive identity view.

However, the disjunctive formulation is susceptible to the charge that it equivocates about the granularity at which mental states and

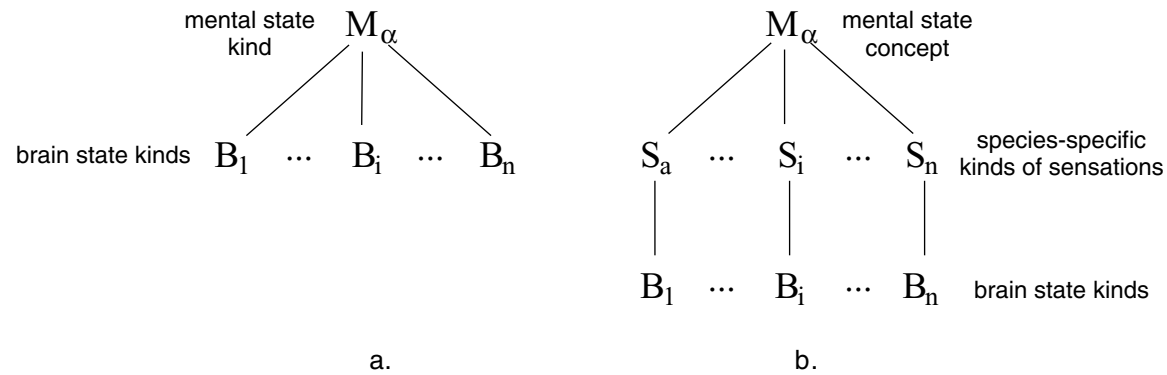


Figure 1.4

Two views of multiple realizability. On the traditional view of multiple realizability (a), a mental state kind, M_α , is realized by various brain state kinds, B_1, \dots, B_n . On Lewis's view (b), a single mental state kind concept, say, *pain*, is realized by different kinds of sensations, S_1, \dots, S_n , in different creatures: pain in humans, pain in octopuses, pain in Martians, etc. Each of these sensation kinds falls under the single mental state concept (M_α), and each sensation kind (S_1, \dots, S_n) is identical to a species-specific brain state kind (B_1, \dots, B_n).

physical states are individuated. If the realizers, considered at the appropriate grain, are diverse, then it is a mystery why, say, *X in humans* and *X in dogs* should be counted as the same kind of mental state. A better way of construing Lewis's intuition is to say that *X in humans* and *X in dogs* are fine-grained distinctions within a coarse-grained mental state kind *X*. Thus, as Kim and Adams urge, they share some coarse-grained property or properties after all, despite being diverse fine-grained states (see figure 1.4).

An alternative is to hold that the fine-grained sensation *X in humans* is different from *X in dogs* or *X in Martians*, and that each is identical to a distinct type of physical state in the respective kinds of creatures. That is, species-specificity is both psychological and biological. This is the picture suggested by Kim (1972, 1998), Flanagan (1992), and Bechtel and Mundale (1999), and I contend that it is the sort of species-specificity supported by Putnam's basic multiple realizability intuition. We think that there could be conscious aliens that differ from us not only in their biology, but in the kinds of mental states they have: "If physical realizations of psychological properties are a 'wildly heterogeneous' and 'unsystematic' lot, psychological theory itself must be realized by an equally heterogeneous and unsystematic lot of physical theories" (Kim 1993: 328–339).

The trouble with species-specificity, as I admit from the outset, is that it sacrifices generality. How, then, do we explain what *X in humans* and *X in dogs* have in common? According to the view now being considered (that psychological states are themselves species-specific), the answer is that they do not have anything in common. *X in humans*, *X in dogs*, and *X in Martians* are not psychologically alike; they are as psychologically and biologically distinct as any three types of sensation. Perhaps we group them because they are typically caused by similar stimuli. Resolving this tension appears to be Lewis's motivation in "Mad Pain and Martian Pain" (1980). Both Lewis and Kim turn to a kind of functionalism to explain what *pain in humans* and *pain in dogs* have in common. For this reason Marian David (1997) dubs Kim's view "conceptualized functionalism."

If we accept species-specificity of this second sort, then we must admit that psychology is not completely general. There could be minded conscious beings to whom psychology does not apply. But

this should not particularly bother anyone: “If Fodor and Putnam were committed to a science of all possible perception, there is no reason to believe they were right” (Block 1997: 129). Likewise, if there are life-forms that are not carbon-based, then perhaps terrestrial biology is not perfectly general. No surprise.

It might be thought that denying that psychology is perfectly general opens the door to additional worries, such as those behind Donald Davidson’s (1970) anomalous monism. Davidson denies that mental states are nomologically regular, that is, that there are general psychophysical laws. But with any identity theory there are lawlike psychological correlations and lawlike psychophysical correlations, both of which Davidson denies.²⁵ Moreover, the generalizations of the species-specific identity theory are universal in scope; the mental states that figure in the theory are just not as frequent as some have thought. So Davidson’s concerns are not sparked by species-specific psychological kinds, nor do they assist multiple realizability arguments against identity theory. Notice, in fact, that anomalous monism is no friend to multiple realizability. Philosophers have tended to focus on Davidson’s denial of lawlike psychophysical regularities. That is the aspect that makes his positive suggestion, anomalous monism, tantalizing; and it has seemed to rule out the possibility that psychological kinds could be “reduced to” neurobiological kinds. But Davidson also denies that there are lawlike psychological generalizations. The multiple realizability argument doesn’t even get off the ground unless there are genuine psychological kinds that might have diverse realizers. If there are no lawlike psychological generalizations, then what is the basis for psychological kinds? If there are no psychological kinds, then there are no multiply realized psychological kinds.

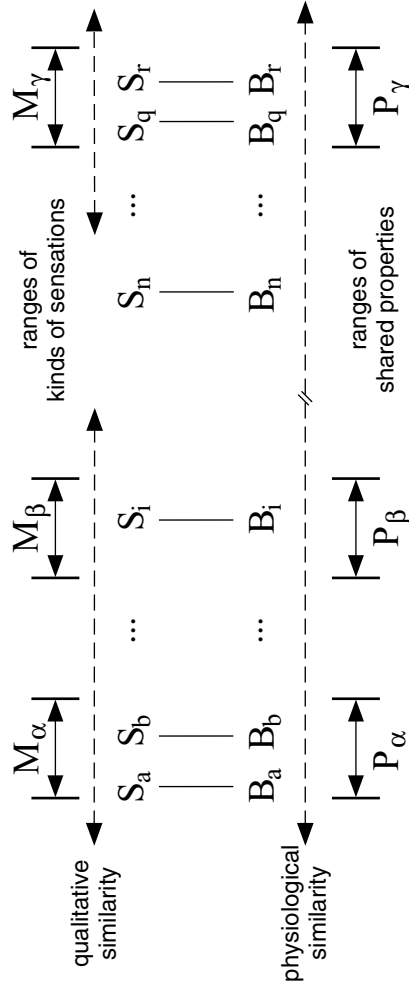
6 Multiple Realizability for Identity Theory

I have argued that Putnam’s intuition that mental states are multiply realizable does not show that identity theory is known to be false or that it makes improbably strong empirical claims. We need only to distinguish the different intuitions that might pass for intuitions of multiple realizability. Standard MR and radical MR are themselves

implausible. Even so, they do not rule out identity claims because, as Kim and Adams argue, the purportedly various realizing systems may share properties. Within individuals and between individuals of the same or similar species, it is quite likely that realizers have the same or similar properties. Thus cases of weak MR and SETI MR can be accommodated within identity theory. And the evidence does not warrant supporting the stronger claims of standard MR and radical MR. It tends to suggest, rather, that kinds of psychological states are themselves as species-specific as their realizers. When creatures are quite different from us physiologically, we may think that they have some conscious states or other but we do not suppose that they have empathetically the same kinds of conscious states that we have.

The view of multiple realizability I am advocating is illustrated in figure 1.5. Particular kinds of sensations, S_1, \dots, S_r , are identical to particular kinds of brain states, B_1, \dots, B_r . Sensation kinds may cluster into coarser, more general species-specific mental state kinds (e.g., M_α or M_β), but insofar as they do, we expect that their members will share physical properties (e.g., P_α).²⁶ Creatures that are similar physically, those that can have P_α or P_β for example, may also have relatively similar mental state kinds, M_α or M_β , say. We should expect human beings and higher primates to have similar conscious mental states because their brains are quite similar to our own. And we should expect the experiences of octopi or aliens to be different from ours to the extent that their brains are quite different from our own.

This sort of species-specific identity has sometimes been thought to block the “reduction” of psychological theories to neurobiological theories, traditionally understood as the translation of psychological theories into neurobiological theories. If psychology can be “reduced” to neuroscience then this would show that identity theory is true by showing that the terms that name psychological kinds in fact refer to neuroscientific kinds. But if psychological kinds are multiply realized, the argument goes, then they do not pick out homogeneous neurobiological kinds and theory “reduction” is not possible, leaving us no reason to assert mind–brain identities.²⁷ To this worry a number of responses are available. One might argue, as Paul Churchland (1982) has, that although biological realizers are

**Figure 1.5**

Multiple realizability for identity theories: an identity theorist's view of multiple realizability. Similar kinds of sensations are identical to similar kinds of brain states. Very different sensations are identical to very different brain states.

diverse, it remains possible that they are alike at subbiological levels. Jaegwon Kim (1989) and Berent Enç (1983) have argued that the local “reductions” that would be licensed by species-specific identities are all that we get in other classic cases of theory “reduction” in the physical sciences, so they are all we should expect for psychology. Another response is to argue, as John Bickle (1998, 2003) has, that the multiple realizability argument depends on an outdated notion of “reduction” that should be replaced by a “new wave” alternative focusing on the relations between theories rather than the relations between ontological levels.

But the identity theorist is not forced to deploy any of those strategies. Belief in psycho–neural “reduction” is one reason—admittedly, a common one—to embrace identity theory, but it is not the only reason. Bechtel and McCauley, for example, press for what they call a “heuristic identity theory,” according to which mind–brain identities are posits that neuroscientists make in the course of formulating theories: “Emphasizing the thoroughly hypothetical character of all identity claims in science, [heuristic identity theory] focuses on the way that proposed identifications of psychological and neural processes and structures contribute to the integration and improvement of our neurobiological and psychological knowledge” (1999: 71). On this view, the search for mind–brain identities is a regulative ideal in the neurosciences. But the heuristic aspect of Bechtel and McCauley’s proposal threatens to undermine the support it may give identity theory. A functionalist might be happy to admit that the practices of neuroscientists involve interim hypotheses of structural identity, certainly within similar species. Nevertheless, the functionalist may suppose, heuristic identities are, in the course of theory building, replaced by more sophisticated realization relations.

Bechtel and McCauley are correct that the justification of identity claims does not await seeing whether completed psychology “reduces” to completed neuroscience. Likewise, John Perry (2001) adopts an identity theory that does not depend on either the success or promise of explanatory or theoretical “reduction” as part of his “antecedent physicalism.” On my view, identities are not intertheoretical devices, but rather figure in the theories themselves. They do not have to be heuristic or provisional or antecedent; they may figure as inductive conclusions, the result of the usual practices of infer-

ence to the best explanation. Lewis (1970, 1972) argues that scientific identities are implied by the theories that posit them. Of course, Lewis believes that theories entail identities by giving the meanings of the theoretical terms and thus fixing the nominal essences of their objects. But Lewis's is not the only model of theories that implies identities. This is also the case of theories that pick out their objects by their real essences; in this case, theories work by fixing the reference of their terms. Either way, identities are what we expect from other sciences; and we should expect mind–brain identities as well. Since most scientific identities are species-specific, we should expect mind–brain identities to be localized. If it seems that we will not find the identities we are looking for, then we will face a hard choice. We might give up our identity claims; but we might decide to adjust the way we individuate the kinds involved in order to permit identities. The latter will be quite likely if we are only trying to accommodate a few resistant cases.

7 Eccentric Minds

On my picture, some conscious creatures could be different from us to such a degree that we must say that they have kinds of mental states only remotely similar to those that we have, even when mental states are coarsely individuated. Creatures that have very different physiological properties (say, those that have P_i may have sensations $[S_q, \dots, S_r]$ but they will be of wholly different kinds, like M_j ; see figure 1.5). The identity theory does not rule out such cases. What this suggests is that at some point the Kim–Adams reply must give way to species-specificity of the sort described by Kim, and by Bechtel and his collaborators. To accept this sort of species-specificity is to accept the possibility of what we might call *eccentric* mental states that are of physically and qualitatively different kinds than our own.²⁸ The possibility that some creatures have such eccentric conscious states is entirely compatible with identity theory. Their eccentric mental states are identical to their eccentric physical states just as our mental states are identical to states of our brains.

Eccentric experiences raise obvious epistemic and methodological problems for how we might detect that these creatures are conscious, or be justified in attributing mental states to them. When creatures

are physiologically like us, we can attribute to them sensations like those that we have. Insofar as a creature is not like us, it does not have those sensations. But now we are considering creatures that are quite unlike us. On what basis do we attribute conscious states to eccentric creatures, and on what basis do we rule out the possibility that rocks and thermostats are not examples of such creatures? The answer is that the identity theorist has available all the tools that any other theory makes use of: functional organization, history, behavior, and the like. Other theories take these characteristics as definitive of mentality and therefore can say with certainty of any creature that it is or is not conscious, or better, to what degree it is conscious. But if we demand certainty, we cannot have it; for an identity theory takes these standard characteristics as only defeasible evidence. The identity theory may have to concede that there could be creatures about which we cannot know with certainty whether they are conscious. But that does not mean that we have nothing at all to say.

Conscious states are related to one another; they have structure. The sensation of yellow is both similar to and different from the sensation of green in ways that neither is related to the taste of Château Haut-Brion. Might there be eccentric colors, colors not related by similarity to the colors we experience? What would it mean for there to be a color that was not related to yellow and green? We do not know what to say (Hardin 1988).

The view I am advocating is not Colin McGinn's (1991) view, which Flanagan (1991, 1992) dubs "new mysterianism," according to which we are not cognitively equipped to understand consciousness. Rather, we are simply not situated to know with certainty whether some creatures are conscious, in much the way that physics holds that we are not situated to know what is going on outside our light cone. There is nothing mysterious about either of these limits. (In fact, there is something suspicious about a theory that has no boundaries.) Some might think it is a conceit to believe that our form of consciousness is even this unique, however trivial that distinction. Far greater the hubris of supposing that we just happen to have a characteristic more cosmically ubiquitous than any other we know of. Even the so-called laws of physics change in the limit, when things are very big, or very small, or very fast. It should not be troubling

that mental categories—those of human psychology—break down when we try to apply them to creatures wildly different from ourselves.

8 Why Identity Theory?

I've been making the case that identity theories are consistent with the available neuroscientific evidence. But some will argue that there is no motivation for holding an identity theory. Daniel Dennett, for example, has raised the concern that there is no motivation for adopting such a theory (2000: 382). If that were correct, it could quash any gains I might have made in arguing that identity theory can handle the generality arguments based on multiple realizability.

Identities, of course, are not deductively entailed by the evidence. Empirical experiments will never record more than correlations; and correlation is in some sense a weaker claim than identity. All identities are correlations, but not all correlations are identities. If correlations do all the predictive and explanatory work and are all that we can directly observe, then the additional posit of identity will seem not only explanatorily superfluous but bought on credit of metaphysical faith that can never be repaid. To distinguish identity from a weaker realization claim, on this view, is to insist on a difference that makes no difference.

The most obvious problem with this objection is that identities are not explanatorily empty. Identity dissolves, with brutal simplicity, a number of problems that are otherwise quite recalcitrant. One, mentioned earlier, is the problem of mental causation, which never arises for an identity theory. If mental states are identical to physical states, then mental states ipso facto have the causal powers of physical states. And we shall see that puzzles such as the explanatory gap and zombies readily yield to identity theory. This should not be a controversial claim; it is widely believed that identity theory is a powerful theory that, lamentably, happens to be false. So how it is that mind–brain identities came to be explanatorily suspect is itself a bit obscure.

We typically arrive at identities by the practice of inference to the best explanation. From the point of view of explanatory practice, to

refuse to accept identities on the grounds that there could be some possible creature for whom the correlation fails amounts to no more than a form of skepticism. We can meet it as we would meet the skeptic, demanding that we be given some explanation for why we should doubt *this* or *that* particular identity. My point is not that we should ignore the modal obligations of necessity claims; I take those quite seriously. Rather, when a theory that is supported by the evidence implies identities, we should not suddenly develop an aversion to the metaphysical claims and try to hedge by embracing only correlations. The naturalized metaphysician takes seriously the metaphysical commitments of our explanatory practices.

There is an even tougher response open to the identity theorist: Perhaps we can make all the same predictions on the basis of correlation alone. But prediction and explanation are different matters. So we might say that a correlation-based psychology would explain nothing at all; rather, it would itself stand in need of explanation. There may be good reason to admit psychological kinds that are identified with functional kinds.²⁹ But one cannot claim that identity theory is a more ontologically extravagant theory than every non-identity theory. Identity theory posits one kind of process, event, state, or property; other accounts require multiplication of at least some of those—of processes, events, states, or properties.³⁰

What, then, of the metaphysical credit with which identity theory is purchased? Yes, identity is a stronger relation than correlation. But identity is not a new or unusual kind of relation, and by invoking identity we are not admitting any new relations into our ontology. Moreover, by insisting on mere correlations we must add rogue psychological kinds to the “furniture of the universe” along with an puzzling new relation, *functional realization*. This has the effect of generating a new problem, the problem of explaining why the correlations hold, a problem that identity theory avoids. Functionalists will argue that the mind–brain correlations hold because brains realize minds. Understanding the realization relation is one of the goals of this book, and I will take up the problem directly in chapters 4 and 5. Whatever the realization relation turns out to be, we will want to know why it holds. This is the metaphysical “open question” for the mind–body problem. Adopting the terms that Kim now

uses to discuss supervenience, we should demand an explanation of the realization relation that is “metaphysically deep” (Kim 1998). But this question makes no sense when the relation is identity. Identity holds of necessity; no further explanation is required or available. So the identity theory circumvents a burden that every other account must shoulder, whether naturalistic or nonnaturalistic.

This claim is widely but not universally accepted. David Chalmers (1995, 1996a) complains that identity is brute, and therefore that identity theories fail to explain the mind–body relation. Similarly, Chalmers and Jackson (2001), responding to Block and Stalnaker (1999), argue that “identities are ontologically primitive, but they are not epistemically primitive” (Chalmers and Jackson 2001: 354). As I understand them, Chalmers and Jackson are calling for a conceptual explanation (they say “reductive”) of the identity, namely, an analysis that would allow, for example, claims about conscious states to be deduced *a priori* from claims about physical states alone without use of any additional principles such as identity claims. If this cannot be done then the identity claim is on no surer footing than a property dualist’s psychophysical laws, for *a posteriori* identity claims are epistemologically no different than very strong correlation claims: “this view may preserve the ontological structure of materialism. But the explanatory structure of this materialist view is just like the explanatory structure of property dualism” (*ibid.*). Thus Chalmers and Jackson urge that one must either adopt analytic materialism (Jackson unpublished) or property dualism (Chalmers 1996a). This line of argument leads us into a territory outside the scope of this book. Since I have limited myself to advocating identity theory in favor of other naturalistic theories, I will leave for another day the questions posed by property dualism for naturalism generally. Chalmers’s view, it seems, is that a nonanalytic identity theory is no better off than other varieties of materialism; even if that is correct, I will be satisfied for now to show that such a theory is at least not any worse off.

Nevertheless, identity theory as I construe it resembles a theory that Chalmers and Jackson mention briefly. According to this possible account, mental states are ontologically “reduced” to (i.e., identified with, I assume) physical states, but those identities do not

support “transparent reductive explanation” of conscious-state facts/claims by physical facts/claims (i.e., deduction of the former from the latter without bridge principles). Although there is more to be said, the view that I advocate has those characteristics. (I will come to the second aspect in chapter 6.) About such a theory Chalmers and Jackson offer: “With this sort of theory, as with a property dualist theory, the explanatory gap between physical and phenomenal might be bridged, but it would not be closed” (2001: 356).³¹ Yet I believe that the identity theory fares better against the explanatory gap (Levine 1983) than Chalmers and Jackson suppose. This is because the identity theory can avoid the metaphysical contingencies that trigger the gap reasoning. It is this issue that we now consider.

I have argued that the identity theory can answer multiple realizability concerns. This is remarkable because multiple realizability is the most widely mobilized argument against identity theories and is often thought to be decisive. If I am right, then identity theory can overcome one of the hurdles that was supposed to reveal it to be inadequate. Now we can turn our attention to the second traditional obstacle to identity theories: Saul Kripke’s modal argument against materialism.