8-2012

Defending the Multiple Realization Argument against the Identity Theory

David Barrett
University of Arkansas, Fayetteville

DEFENDING THE MULTIPLE REALIZATION ARGUMENT AGAINST THE IDENTITY THEORY

DEFENDING THE MULTIPLE REALIZATION ARGUMENT AGAINST THE IDENTITY THEORY

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Philosophy

By

David Alan Barrett
Hendrix College
Bachelor of Arts in Philosophy, 2006
University of Arkansas
Master of Arts in Philosophy, 2008

August 2012
University of Arkansas

Abstract

A classic argument in the philosophy of mind is that the identity theory is false because mental state types are multiply realized in brain state types. In this dissertation I provide a detailed elaboration of the argument and a defense of it against a few of its prominent contemporary critics. Finally I offer empirical evidence from inter-species differences in humans and monkeys, and also from a case of extensive neural plasticity, which shows that mental state types are multiply realized in brain state types.

This dissertation is approved for recommendation
to the Graduate Council.

Dissertation Director:

____
Dr. Jack Lyons

Dissertation Committee:

____
Dr. Eric Funkhouser

____
Dr. Thomas Senor

Dissertation Duplication Release

I hereby authorize the University of Arkansas Libraries to duplicate this thesis when needed for research and/or scholarship.

Agreed ________________________________
David Alan Barrett

Refused ________________________________
David Alan Barrett

Acknowledgements

I would especially like to thank my director, Jack Lyons, for his help in completing the present work. Without his advice and patience in reading countless drafts and rewrites, the end product would have been much poorer and much longer in coming. I would also like to thank the other members of my dissertation committee, Eric Funkhouser and Thomas Senor. Whatever I get right probably owes very much to their help. Whatever I got wrong is certainly my own doing.

Table of Contents

Chapter One: What are the Identity Theory and Multiple Realization?

Undoubtedly a classic argument—perhaps *the* classic argument—in the history of the philosophy of mind is to show that identity theory is false because mental states are multiply realized. Beginning chiefly with papers by Place (1956) and Smart (1960), the dualistic outlook on the old mind-body problem was replaced by a thoroughgoing physicalism. In particular these writers argued, mainly via Ockham's Razor, for the view that mental state types and brain state types are identical. Of course, as so often happens in philosophy, there was some general agreement (at least in terms of commitment to physicalism) combined with much criticism. In this case, there was a general agreement about accepting physicalism, but the kind of physicalism to adopt became the main issue. Putnam (1967) and Fodor (1968) and (1974) represent the initial and starkest criticism of the early advocacy of identity theory, arguing that classic line about mental states kinds being multiply realized in various brain state kinds.

The influence of this Putnam/Fodor criticism was enough to make what came to be called nonreductive physicalism the received theory of mind. And, from what I can tell, nothing much has changed in recent years. That is not to say that no contemporary criticism exists of the view—or especially of the multiple realization argument used to get there. Indeed most of the writers who shun nonreductive physicalism have also published explicit rejections of the Putnam/Fodor argument. Examples of such writers include Bechtel and McCauley (1999), Bechtel and Mundale (1999), Boyd (1999), Kim (1992) and (2005), Millikan (1999), Polger (2002) and (2004), Richardson (1979), Sober (1999) and Shapiro (2000) and (2004).

Somewhat surprisingly, amidst what clearly constitutes a substantial and recent backlash to the received view, very few writers have bothered to deal with various criticisms leveled by the writers listed above. The purpose of my dissertation is thus roughly to provide such a response. I do not plan to respond individually to every argument or view contained in the above works, which would just take too long. Rather I would like to spend most of what follows developing this classic multiple realization argument against identity theory to some detail, hoping that such a development and a defense of the way in which I do develop the argument will form a compelling response to the general trends present in these works. Where I think battles against the individual arguments or views are worthwhile, instructional, or just plainly necessary (with respect to the line I want to take), I will save the back third or so of the dissertation to wage them.

But there will be plenty of time and space for the dialectic from Smart to Shapiro. I will cover a good deal of it eventually. First we should do some preliminary work explaining more precisely what the early identity theorists were up to, the arguments from which they drew support for their view, and then finally the details of the Putnam/Fodor argument which came to dominate the discussion. Only after this orientating work will the development of the argument against identity theory and the defense of it I try to provide from more modern criticism be fully comprehensible. It is to that work I turn first.

Section 1.1: Explaining Some of the Key Concepts

As is customary, it is best to start by getting clear on some of the key issues and terms. The obvious starting place is to be clear about what is meant by ‘multiple realization.’ One attempt at explaining it is: a kind, or property, or type, is generally said to be multiply realized if there are *many different ways for that kind* (property, type) *to be*. And, of course, a kind is multiply realizable if it is possible (pick your modality) for it to be multiply realized. There will be much opportunity to discuss the notions of ‘kinds’, ‘properties’, and ‘types’ as we go along. For now, let us pass over those worries and try to get an intuitive grasp of the more basic issue of multiple realization. To that end, think first of a mousetrap. Though ‘mousetrap,’ I take it, does not denote any serious kind picked out by some scientific framework, it will serve as a useful example of how something can possess ‘different ways of being.’ The typical kind of mousetrap is the one with a spring-loaded bar, where, once the trip is touched, the bar snaps down over the area with the bait. But certainly that is not the only way a device could function as a mousetrap. There are, for instance, devices which merely trap the mouse in a cage, or glue-like traps that use an adhesive surface with bait to trap the mouse alive. Indeed, any device that serves to trap the mouse—alive or dead—counts as a mousetrap. Mousetraps do not have to *be* any particular way in order to belong to the kind (or category, or whatever). It is in this sense that we say mousetraps are multiply realizable (and in fact, multiply realiz*ed*, given all of the different devices one can purchase).

Another example one finds in the literature (see, e.g., Shapiro, 2004) as an attempt to explain multiple realization, very close to the above locution about ‘different ways of being’, is *sameness through difference*. The relevant sameness and difference obviously need some spelling out, but the mousetrap example will again give us the gist of the idea. Given the variations of how exactly each kind of mousetraps function—for instance, some kill, some do not—there is clearly some bare differences in each individual device. But there is also an important sameness to keep in mind: viz., they all manage in their different ways to capture mice. The persistence of this sameness in the face of the different functions applied to get the job done accounts for the multiple realization of mousetraps. Their being multiple realized is constituted by the fact such an abstract sameness (the ability to capture mice) is captured by particular mechanisms with differences in their functional properties.

But to put the issue in the most exact terms (the terms by which I make the defense of the 'multiple realization' of psychological state types in neurological state types), a kind counts as multiply realized if there is a *one-to-many mapping* from the higher-level scientific kind to the lower-level scientific kinds (the 'realizers').[1] Setting aside the issue about the realization relationship itself, we can better fasten down an understanding of talk about one-to-many mappings by talking about mental state kinds and brain state kinds. If the multiple realization of mental state kinds were given, adopting this other idiom for explaining the concept, we would say that a particular mental state kind stands in a one-to-many relationship with other brain state kinds. Let us try to make the example more concrete. Take it that a representative mental state kind is 'pain', and take it that there are at least three different kinds of brain states: 'A-fiber stimulation', 'B-fiber stimulation', and 'C-fiber stimulation'. We could thus frame the multiple realization of 'pain' by saying that the kind is realized in these different kinds of brain states ('A-fiber' and so on). Imagine along with Putnam (1967) that octopi and humans have very different brains, though they still both have the same mental states of being in pain. Assuming we find that each species has a different brain state kind correlated with its mental state kind of pain, we would have evidence for 'pain' being multiple realized (in whatever respective brain state kinds we find in octopi and humans which are correlated with pain). I might as well also mention now, since I will be trying to appeal to cases like these later on, that multiple realization might also be

[1] For more commentary on why I think this way of understanding 'multiple realizability' is better than the first two attempts, and to see why it might matter how the phrase is supposed to be understood, see section 4.2 of Chapter 4. Until that time, I do not think any of the work I do will depend on any particular way of understanding multiple realization. In particular, when it comes to individuating the realized and realizer kinds, which I do in Chapters 2 and 3, it will not be important to get clear on this.

found within a single individual over time—not just in sameness of psychology across differences in brains of various species. This seems loosely to be the type of consideration Block and Fodor (1972) were appealing to in defending the multiple realization of mental states, citing the ‘equipotentiality’ of the brain. The idea is that perhaps through recovery from some type of brain damage, or maybe just through aging, our own individual mental states could well be realized by different brain states. After having a stroke, a person might have their A-fiber destroyed without losing the corresponding mental state kind. Instead, that mental state kind is now realized in, say, the subject’s B-fiber through some sort of plastic recovery.

Hopefully the foregoing gives the reader a decent sense of what is meant by ‘multiple realization.’ Let us now contrast that this notion with ‘identity theory.’ As we said earlier, it was writers like Place and Smart who first began arguing for this view. Their main opponent on the mind-body question at the time was actually dualism. Mainly on grounds of simplicity, they argued that a better understanding of the mental would be simply to reduce it to the brain, rather than its being of a different kind of substance altogether from the physical. One can understand the notion of reduction here roughly in the way that ‘light’ was reduced to electromagnetic waves, or ‘warmth’ to high average molecular kinetic energy. In the same way that scientists came to state and defend these theoretical identities—the identities of the entities posited by different conceptual frameworks—identity theorists suggest we understand that mental states kinds are identical to brain state kinds. This, then, is the most straightforward way of advancing the main claim of identity theory: mental state types just *are* brain state types.

So far I have been playing fast and loose with ‘types’ and ‘kinds,’ using them equivalently. Now would be a good time to attempt to clarify these notions, since formulating the different varieties of physicalism will not be possible without doing so, and since the multiple realization argument has been historically very intimately bound up with talk about ‘reduction’. It will not be necessary to be *completely* clear on these notions of kinds and types, but certainly I can say more to make the identity theory in particular as plain as possible and to prepare for discussing theoretical reduction, autonomy, and the like later on.

The notion of ‘type’ is best contrasted with the notion of a ‘token’. If I have, for instance, a large bag of pennies, then we can say that I have a bag with numerous *tokens* of the same *type* of thing. A penny is a type of coin, and each individual penny counts as a token of the type ‘penny’. With respect to the identity theory, we can see the difference that tokens and types make. The identity theory claims the identity of *types* of mental states and brain states, not *just* tokens. The idea is that a mental state type—say, ‘pain’—is identical to some brain state type—say, ‘C-fiber stimulation’. The claim is *not* (merely) that this individual token (of the type) pain is identical to this individual token (of the type) C-fiber stimulation. Rather the identity theorist thinks, roughly, that this type of mental thing is identical to this type of brain thing. This is why the identity theory is sometimes called *type* physicalism, or *type* identity theory—as opposed, of course, to mere *token* physicalism, or *token* identity theory.

To be absolutely clear, mark the logical relations that might hold between a type physicalism and token physicalism. For instance, the former certainly implies the latter: if one supports the idea that mental state types are identical to brain state types, then necessarily one must support the idea that each individual mental state is identical to some other individual brain state. But, importantly, the implication *does not go the other direction*. One can hold the latter without being committed to the former: that is, one may believe that each individual mental state token is identical to some brain state token or other, but think it is false that mental state types are identical to brain state types.

Now, the notion of a 'kind' we can contrast perhaps best with a mere 'property.' Following Fodor (1974) here, we can draw this difference by noticing that every science uses theoretical and observational predicates in order to subsume events in the world under its laws. For instance, geology might use phrases like 'mountains' in order to articulate some of the laws that make it the science it is. So, too, economics can articulate its laws by talking about things like 'monetary exchange'. It is properties like these which are picked out by the so-called *natural kind* predicates of a science—the predicates which taxonomize events in the world under a particular description. And, again, these natural kind predicates are the basic building blocks of the laws of some science. The hope is that, by using the predicates of some science, by slicing up the world in terms of 'mountains', 'flowers', 'water', or 'C-fiber activity', we can express true generalizations about how the world works. Roughly speaking at least, that seems to be the entire point of a science. But Fodor also reminds us that there are many other true descriptions of a given event which do not employ the vocabulary of any science. His example is that there are presumably many events which consist of something being transported within three miles of the Eiffel Tower. Surely there is no science, however, which employs as a part of its vocabulary—designates as a natural kind—the predicate 'is transported within three miles of the Eiffel Tower'. This shows that such a *property* does not constitute a natural *kind*. The difference between the former and the latter can be put accordingly: of course every kind is a property, but not every property determines a kind. So, with respect to the mind-body question, we can take it that the identity theory is supposing an identity of kinds. That is, when we claim that mental state types are identical to brain state types, the relevant notion of a 'type' is a kind, as picked out by some science or other. In our case, as we will see later on, the sciences will be psychology and neuroscience. For now, though, I only want to make it clear that in what follows I use 'kinds' and 'types' interchangeably.

Section 1.2: Functionalism, Identity Theory, and The Argument From Multiple Realization

Precisely what the identity theorist claims ought to be as clear as possible now. Of course the state of the sciences involved may not yet be mature enough to identify the kinds—hence the ubiquitous and utterly fictitious example of 'pain = c-fiber stimulation'—but that does not stop us from precisely articulating the view: *mental state kinds are identical to brain state kinds*. If a creature ever instantiates a particular mental state type, then that creature must instantiate the relevant brain state type, and vice versa. Next we should look at the criticism of this view, so we can make just as clear exactly how the multiple realization argument is supposed to work against it. As I mentioned above, the chief objectors were Putnam and Fodor. To understand their criticism best, we need to talk about the view of mind they espoused as the correct alternative: functionalism.

There are different kinds of functionalism one could adopt—e.g., machine state functionalism or computationalist functionalism. For our purposes, though, it is okay to overlook these distinctions. What these positions all have in common is the idea that we should individuate mental states kinds by their causal roles. That is, a particular mental state is of a mental state *type* in virtue of satisfying some causal role. Just to provide a concrete example, some functionalists might think that mental state types are individuated by their causal connections to environmental stimuli, other mental states, and to behavioral output. So, we can classify 'pains' as those things that are caused by kicks, pokes, and scratches (etc.), cause other mental states like anguish, fear, and practical reasoning designed to stop the noxious stimulus (etc.), and also causes us to cry, wince, or say things like 'Ow!' (etc.). The claim, then, is that the defining feature of a mental state type, what makes it the type of mental state that it is, is the causal role it plays within the mental system of which it is a part (rather than its being, e.g., identical to brain state types). As the functionalist slogan goes: it is what it is because it does what it does. We can classify our mental states by virtue of their function, or what they do.

Note the interesting consequence here for the real ontological question: one *could* be a functionalist and maintain dualism! So long as the nonphysical soul stuff, or whatever, satisfies the causal role by which we type a given mental state, a dualist ontology is compatible with a functionally specified notion of mentality. Why is this important? Because it highlights a point which is critical for understanding the early objection to the identity theory: for functionalists, *we can individuate mental states without worrying about what they are made of*. Compare the identity theory: mental states are only made of brain stuff; there is no room for dualism (which makes complete sense because identity theory was founded, so to say, in opposition to dualism). And why care about the ambivalence of the 'realizers' of a mental state? Because this lack of concern for the 'realizers', at the very least, allows for the conceptual possibility of mental states being multiply realized. Again, compare the identity theory: according to the typing of mental states by brain states, a mental state cannot be realized by anything but brain states. The apparent dissonance between multiple realization and the identity theory should thus start to make sense. Prima facie it seems that if mental states can be realized in states as diverse as brains, computer-like entities, or in ghost-like soul substance, the identity theory is in big trouble, claiming as it does that one only finds mentality when brains are also around.

Something like this line, in fact, was the dominant reason for the rejection of the identity theory. Putnam (1967) was the first to articulate the worry. He assumed, plausibly at the time, that various species could share the same mental state without having anything importantly in common in terms of brains. That is, it seemed plausible that, say, octopi experience pain in just the same way that humans do. But, on the other hand, it seems implausible to suppose that octopi and humans have very similar brains—or, more specifically, that the brain states underlying the pain through the two species are very much alike. Hence it seems reasonable to suppose that pain is multiply realized in humans and octopi. Given the identity theorist's commitment to the idea that any mental similarity is realized by similar brain states, the view thus runs afoul of what strikes us as empirically probable. Of course, the functionalist does have a view which is compatible with these facts about octopi and humans: let the brains be as dissimilar as you like, so long as they both realize the appropriate causal connections the relevant mental similarity will be preserved.

For my purposes, I do not care about arguing for the merits of functionalism over the identity theory. I would rather stop at the problems multiple realization poses for identity theory and not bother to worry about putting forth a positive theory of mind. Later I will adopt a computationalist notion of a psychological state, but my reasons for this have nothing to do with any ontological preference. Officially I remain neutral here about the truth of functionalism. But instead of complicating things now at such an introductory stage, it would be best simply to lay out the main argument I am defending. Having made clear what the identity theorist is claiming, what multiple realization amounts to, and also where the functionalist fits in to this discussion, we are set to do so. I will not bother to make the argument formally valid, however. The version put forward here is intuitive, easy to remember, and perfect as a magnet for the criticism I want to present now and evidence I want to present later.

Premise 1: *Mental state types are multiply realized in brain state types (i.e., mental state types are not in a one-to-one correspondence with brain state types).*

Conclusion: *So the identity theory is false*.

Note that, as it is shown, I am concerned with mental state types *actually being multiply realized*. The evidence I will invoke later on is intended to show *not* that such kinds are multiply realiz*able*, but that in real organisms they are in fact so realized. I take it, without any argumentation, that the identity theorist is committed to the idea that the relevant identities are metaphysically necessary. 'Pain', as far as I can tell, rigidly designates pains, and so does 'C-fiber activation.' 'Pain' (metaphysically) could not have referred to anything else other than those unpleasant experiences we have, and 'c-fiber activation' could not have referred to, say, the spilling of a cup of water. Hence their being identical is a metaphysical necessity; any metaphysically possible world in which you have the one, you must have the other. If that is true, then merely showing that mental state kinds can (metaphysically) possibly be instantiated in various brain state kinds would be enough to show the identity theory is false. I am not interested in making such an argument. Instead, I want to make the stronger argument that, in fact, mental state kinds are multiply realized in brain state kinds. Thus the bulk of my dissertation will be a defense of the premise above—either by offering direct evidence in its favor or by defending it from various forms of criticism.

Though I have tried to remain steady in my usage, below I might run back and forth between the two phrases 'realizability' and 'realization' somewhat carelessly. In what follows, I ask the reader to remember my emphasis on multiply realized (psychological/mental) kinds and overlook any use of multiple realizable (psychological/mental) kinds, since strictly speaking I am not interested in them. I think this would have been obvious without a warning, given everything else I say, but it is probably best to be clear up front.

With the basics of the argument out of the way, the next thing to do is survey the criticism the argument has received. The main reason I am attracted to defending this argument is because of that criticism, and of course because I have never thought the criticism was any good. In order to evaluate the defense I make of the argument, and to assess the evidence I present in favor of it, it will be important to respond to the writers who have rejected it. So I will take the time to discuss some of the more famous material. I will not provide a thorough exegesis of anyone just yet, though. For now what is important is placing the argument against identity theory in the context of the rest of the literature. To that end, there is no great urgency to undertake serious scholarly work; merely making clear the contemporary ways of resisting the argument is good enough. Sometimes understanding the ways to attack an argument is also beneficial for understanding the argument itself more fully. Surveying the criticism will also make the direction and content of the dissertation more understandable. At any rate, when, after developing the argument the way I want to, it becomes important to focus on an individual critic, I can do so with more precision. Attempting to elucidate everything that matters at this point probably will only confuse things, forcing the reader to keep too many moves and claims in mind that are best delayed until later.

Section 1.3: Popular Criticisms of the Multiple Realization Argument

As with any argument, there are typically two ways of mounting a challenge: either reject the premise(s), or try to show that the reasoning is no good (that the argument is invalid). The literature is crowded with those writers who have done both. In other words, there are many (probably most) who have simply denied that mental states are multiply realized—to show that the premise is false. And there are those who have tried to show that the multiple realization of mental states is compatible with the identity theory—to show that the argument above is invalid. For now I want to look at some of the important contemporary criticism of the argument that has attempted either of both of these moves.

First we can consider Thomas Polger (especially, 2002 and 2004). He can be classed in the group of critics who attempt to block the inference from the multiple realization of mental states to the falsity of identity theory. He does so first by providing, as it were, a taxonomy of multiple realizations. He notes that Putnam's original intuition was merely that we should not anticipate that every creature with roughly our psychology would also have a brain with states like ours. But this initially plausible claim, with which even Polger agrees, has been stretched too far in recent discussions of multiple realization. He identifies four different cases along the 'spectrum':

i. At least some creatures that are not exactly like us in their physical composition can be conscious. (Weak-MR)
ii. Some creatures that are significantly different from us in their physical composition can be conscious. (SETI-MR)[2]
iii. Systems of indefinitely (perhaps infinitely) many physical compositions can be conscious. (Standard-MR)
iv. Any (every) suitably organized system, regardless of its physical composition, can be conscious. (Radical-MR)

[2] 'SETI' is short for 'search for extraterrestrial intelligence'. In short, this construal of multiple realizability is the one that takes seriously the idea that Martians, with their silicon-based brains, could share our psychology.

What he wants to do with this classification in hand is to argue that the latter two kinds of multiple realization are either flatly implausible or the artifact of a previous commitment to functionalism—which would make an argument proceeding from them question-begging. The former two, on the other hand, are plausible, but the identity theory can accommodate their truth. It is in this respect that Polger appears to deny the move from multiple realization—suitably spelled out—to the falsity of identity theory.

Some of the details go as follows. Standard-MR and Radical-MR are clearly pretty strong notions of multiple realization. They are strong in the sense that, compared to cases of mere neural plasticity, holding that, say, a collection of gum and paperclips could realize human psychology is a very striking portrayal of the multiple realization of mental states. Now, according to Polger what makes one inclined to accept something like Standard-MR or Radical-MR is probably a prior commitment to functionalism. That is, the plausibility of either of the more extreme kinds of multiple realization is held hostage to a functionalist theory of mind. If functionalism is the right view of mind, then it does seem rather as if (at least) Standard-MR is likely. But the argument, of course, is supposed to settle on a conclusion about the theory of mind. It is no good, says Polger, sneaking in theory of mind from the beginning. If our goal is to make an argument to rule out a particular view on the nature of the mind, we had better not be using a premise which tacitly presumes such a position—or else the argument merely begs the question at issue. So, our worry thus boils down to how plausible either Standard-MR or Radical-MR is *independently* of our finding functionalism plausible. Polger maintains that no one, surely, finds either of those brands of multiple realization to be plausible absent a predilection for functionalism. As Polger puts it, functionalism and those two varieties of multiple realization just go together.

What about the weaker versions of multiple realization? Here Polger does not think there is any incompatibility with identity theory. He explains that whether we classify two states or entities as belonging to the same kind or type is independent of whether every single property of the two states or entities is shared. That is, two things could vary wildly vis-à-vis the sum total of their properties, but so long as they share the properties which are individuative of some kind these two things will be classified together. This was the point we made about mousetraps. Though two individual mousetraps can be composed of vastly different stuff, but they still count as mousetraps because they share the same all-important (type-individuating) property of being able to trap mice. The suggestion is that Weak-MR and SETI-MR leave it open that the properties by which we classify brain states might establish a *sameness* of brain states between us and other creatures even though these creatures are "not exactly like us" or even "significantly different from us" in their physical composition. In this way, we can state Weak-MR or SETI-MR as our premise about the multiple realization of mental states, from which we would not be able to conclude the falsity of identity theory. The bare fact that these other creatures are physically different from us—even significantly so—does not yet establish that those differences are relevant for demonstrating a difference of brain states. So Polger, in dividing up the different notions of multiple realization as he does, attacks the multiple realization argument in two ways: either we deny the premise (when we construe multiple realization as Standard- or Radical-MR), or we deny the validity of the inference (when we construe multiple realization as Weak- or SETI-MR).

This kind of line leads perfectly into a discussion of Lawrence Shapiro's objections to the multiple realization argument (see, in particular, his 2000 and 2004). In general, he does not deny the move from multiple realization to the falsehood of identity theory; rather he denies the premise. The way in which he tries to show that mental states are not multiply realized is roughly by exploiting the line Polger was only hinting at above. In other words, Shapiro exploits that idea that advocates of multiple realization have to show that some mental state is, in fact, realized in *different* brains state types. You have to put the 'multiple' in multiple realization. For instance, it does not seem quite right to say that two individuals of some type (say, mousetraps) which differ only in color count as *multiple* realizations of that type. So clearly some work needs to be done in order to say what makes some other individual a relevantly *different* realization. Shapiro's attack begins here.

What he offers is a criterion for establishing that two realizations are actually different. He says that a kind counts as multiply realizable if there are different ways to bring about the function that defines the kind or type in question. If the only difference between two individuals is some property which does *not* contribute to the performance of the function (which types them as the same), then they do not count as *multiple* realizations of that type.

To make this more concrete, consider Shapiro's example about corkscrews. Of course we count things as 'corkscrews' merely by their ability to remove corks. And any old thing which can perform that general function gets to count as a member of the 'corkscrew' group. There are also, of course, many different ways of getting that particular job done. There are, for example, waiter's corkscrews and winged corkscrews. The former uses a lever to pry the cork out of the bottle, while the latter uses a rack and pinion to perform the same function. It is in virtue of performing the same function in different ways that constitutes the multiple realization of corkscrews. On the other hand, if we have two winged corkscrews that differ only in their color (one is red the other black), then these two corkscrews do not count as multiple realizations of the type. Here, rather, we only have a difference in a property that is not causally relevant to performing the function (which individuates the type—e.g., which makes it a corkscrew). Similarly, and more startlingly, Shapiro also thinks that two waiter's corkscrews made of different material—say, aluminum and steel—would not necessarily count as distinct realizations. The point, again, is that neither of the materials contributes to the causally relevant properties which are doing the work. Both the aluminum corkscrew and the steel corkscrew are *rigid*, of course. And it is the rigidity, at bottom, that is doing the functional work (of getting the cork out of the bottle). The aluminum and the steel, however, do not contribute to the function of removing the cork. In this case, they are just incidental properties of the corkscrew, like the color. Hence, for Shapiro, these corkscrews made of different metals do not count as different realizations of the type 'corkscrew'.

Armed with this particular understanding of multiple realizations, Shapiro (2004) goes on to argue that there are certain constraints on brains (which can realize our particular human psychology) such that it is empirically unlikely that we will find any different realizations of our psychological states. For instance, take our visual abilities. Surely approximating very nearly our visual acuity is necessary for having the same visuals states that we have. Well, given that the density and organization of our visual reception correlates with our acuity, it seems like a constraint on any realization of our visual states that it must contain the rough number and organization of visual receptors. Or, to use another example, it is well documented that lateral inhibition is necessary in order to heighten contrast. That is, the cells responsible for conveying retinal information must be connected in such a way that adjacent cells have an inhibitory effect on one another. Without going too deeply into the details, the general point is that any realization of our visual abilities has got to have something like lateral inhibition in the works. So, assuming that these are bona fide constraints on the actual processing of our visual information (on the properties causally relevant for producing our visual states), we can say in advance what sorts of brains a creature will have who shares our visual states. In effect, these constraints rule out the possibility that there could be *multiple* realizations of our mental states. One might think, of course, that it is possible to get silicone chips to process information via the use of lateral inhibition—and so it follows that our mental state types are multiply realized. But do not forget Shapiro's remarks about what count as *multiple* realizations: the physical constitution of the functional kind in question is irrelevant if that constitution does not make a functional difference. If the silicone chips process the visual information via lateral inhibition (along with all the other constraints imposed on generating our visual acuity), then we have got the exact same function used to see (i.e., provide a certain level of visual acuity). And that means we do not have *different* realizations of our visual abilities. So we do not have multiple realizations of those visual states either, even if we have got a 'brain' made out of silicone chips. *That* is why it is important to Shapiro to discover constraints on brains.

There is one way generally to attack multiple realization claims: show that things are not so different at the level of 'realizers' (the brain state types are not different, so that there is still a one-to-one correspondence between mental state types and brain state types). Shapiro, in particular, reminds us of this line of attack. There is, it should seem obvious, another general attack to make. Rather than maintaining that the brain states (or whatever is at what I am loosely calling at the moment the 'realizer' level) are not of different types, you may also try to claim that the mental states are not of the *same* types. That would amount to saying that you have a many-to-one, or many-to-many, correspondence of mental states types to brain state types, instead of the required *one-to-many* arrangement (i.e., to be clear, one mental state type to many brain state type arrangement). Though Shapiro is also fond of this other line of attack, he is not the only one who pursues it. Polger, as we have already seen, is another who explicitly uses this tactic to deny multiple realization.

This other line, to make it less abstract, has some intuitive appeal. We do, in a loose sense, think that octopi, dogs, and all the rest feel pain like we do. But how many of us are sure that they feel *just the same* pain that we do? If we decide that a very strict phenomenal dimension is relevant to typing mental states—that is, we think pains are of the same type in virtue of, inter alia, *feeling* to a fairly strict degree the same—then it becomes prima facie plausible that octopi, dogs, and us do not all occupy the same mental state type (pain, or whatever). So notice the problem raised for multiple realization: if dogs do not feel just the same kind of pain we do, then we cannot say that that particular mental state type is multiply realized. Rather, it seems like what Kim (1992, 2005) claims about species-specific reductions is likely to be true. Our pain is realized by a particular brain state type, and a dog's pain—which is of a different kind, because it feels different—is realized in some other brain state type. At that point, it does not even matter whether those brain state types are the same or not.

Of course we are free not to type our mental states by reference to sensations, or phenomenal feels. But that gets ahead of the point I want to make at the moment, which is simply that multiple realization demands—in addition to bona fide difference of brain state types—*sameness* of type of mental states. Critics are free to attack in either direction. Shapiro and Polger are examples of writers who have. It is also worth pointing out that these are both ways of denying the premise in our argument above.

Last, we should rehearse a different kind of criticism. William Bechtel and his collaborators (see especially Bechtel and McCauley, 1999, and Bechtel and Mundale, 1999) have preferred to look at empirical work in neuroscience rather than focus on more metaphysical issues (e.g., what counts as a *multiple* realization of some kind, or whether multiply realized properties are actually kinds, etc.). Of course many agree that the evidence for multiple realization is empirical in nature; Bechtel and company are unique in arguing that, when we look at the work of neuroscientists, we see that success in their field depends on the truth of identity theory.

The idea is fairly simple. Supposing that multiple realization of psychological types were the case, it would make very little sense to use advances in our understanding of, say, visual processing in the macaque monkey as a model for visual processing in humans. Yet, contrary to Putnam's intuition, this is precisely the sort of method neuroscientists employ in order to understand how humans process visual information. Perhaps philosophers would be surprised to learn that the vast majority of research that is done in order to find out about human psychology is carried out on members of other species. Now, these sorts of experiments are not typically done on just *any* kind of animal—i.e., you do not see researchers into human audition look into auditory processing in bats. But it is certainly clear that research on the psychological abilities of monkeys, rats, and cats, to name just a few kinds of animals, has shed some light on human psychology. Think also of the continued use of positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) as methods for localizing cognitive function in the brain. These tools would be worthless for neuroscientists and neurosurgeons if there was not a fairly substantial similarity of brains across brains or even across individuals. In fact, fMRI and PET scans are probably the most important tools used by cognitive neuroscientists in coupling cognitive abilities and brain areas. Simply scanning the titles of papers in neuroscientific journals makes their importance obvious.

Another way of putting the point is that you do not see the taxonomies of psychologists and neuroscientists evolving independently of one another. One of Fodor's main conclusions was that the multiple realization of special science kinds ensures the methodological *autonomy* of the special sciences. A science is autonomous from another science when, roughly speaking, one can fruitfully do the science in question largely in ignorance of laws and entities of the other science. Take psychology for example. Its autonomy from neuroscience, if it is an autonomous discipline, would amount to the idea that one could do serious psychological theorizing and research while knowing next to nothing about the brain. On the other hand, if we take it that psychological kinds are in a one-to-one correspondence with neuroscientific kinds, then really all psychological laws and theories are simply notational variants on neuroscientific laws and theories. One would only have the same laws couched merely in another vocabulary. And proceeding with psychological research without knowing anything about the brain would be foolhardy to say the least. From Bechtel's and Mundale's perspective, psychological classifying and neuroscientific classifying have largely started to intermingle in just the way that advocates of multiple realization would have thought unlikely. Their main example is human visual processing. Work by those like Hubel and Wiesel (1962), Ungerleider and Mishkin (1982), and van Essen, Anderson, and Felleman (1992) have contributed positively to formation of psychological theories of vision. Of course, contributions can run in the reverse as well: psychological theory confirmation might lead us to take seriously a particular neuroscientific theory. Anyway, the point is simply that such a co-evolution of taxonomies shows that researchers in either discipline assume a heuristic identity theory. In particular, they must be assuming that, say, a lesion in a particular area of the brain will correlate with a very specific kind of cognitive deficit—a finding that would seem unlikely if cognitive state types are actually realized in many different kinds of brain states. Going the other way, these researchers might also assume that the presence of a kind of cognitive ability implies the presence of particular brain structure—again, to the surprise of those who claim there is a one-to-many mapping of cognitive state types and brain state types. That psychologists and neuroscientists have been able to influence one another's work—vis-à-vis taxonomies and theory formation—indicates that there is substantial sameness of correlation of particular psychological abilities and particular brain state types across species, all of which flies in the face of Putnam's old intuition about pain and octopi.

What Bechtel and Mundale offer as a diagnosis for why philosophers have found multiple realization so tempting is also important. They assert that philosophers have tended to equivocate on the 'level of grain' at which they specify the relevant mental and brain state kinds. In particular, philosophers have taken a fairly liberal view of same mental state type (i.e., they have operated with a fairly *coarse* grained view of mental state taxonomy). For example, as we remarked above, it seems plausible that there are differences in the way pains *feel* for members of different species; they are not exactly the same then. Nevertheless, when discussing the merits of multiple realization, philosophers have been willing to lump those different pains into the same kind. On the other hand, philosophers have tended to operate with a rather *fine* grained view of brain states. That is, just about any subtle difference in the brains which are compared is enough to justify lumping either brain (state) type into a different category. Well, if we were to equivocate across the level of grain at which we were classifying the relevant types, then we can easily, basically trivially, find a one-to-many mapping of mental state types onto brain state types.

The only legitimate way, however, of establishing this one-to-many correspondence is to maintain the level of grain *constant* across the classifying of the types. Bechtel and Mundale do not particularly think it matters what level of grain we adopt, though. So long as we remain constant, so they say, multiple realization will not seem plausible. If we adopt a coarse level of grain, then we will not actually find that the brain states are of a different kind (we only get a real difference of kind if we adopt a really fine level of grain). And if we adopt a fine grained view, then we will not have the requisite sameness of mental state types (which we can only get when adopting a rather coarse level of grain). Either way we go, then, we end up having no support for the view that the *same* mental state kind is multiply realized in many *different* kinds of brain states. The problem is simply overcoming the temptation to make the subtle equivocation.

Section 1.4: An Outline of the Rest of the Dissertation

With the discussion of Bechtel and the rest out of the way, the layout of the literature against Putnam's and Fodor's famous view should be substantially clearer. What would help now would be to use some of the general worries, drawn from the work of the above writers, to explain in rough outline the direction I plan to take the rest of the dissertation. In that way, I can simultaneously make a bit of a roadmap for the reader and show how that direction I plan to take things fits into the existing literature.

First of all, there are in general two sorts of avenues I have to explore: conceptual and empirical. On the conceptual side of things, I hope it has struck the reader as obvious that I will *have* to say something about mental state kinds and brain state kinds—viz., how to individuate them. Recall Shapiro's and Polger's attacks: they point out that we need genuine heterogeneity at the realizer property level (the level of brain state kinds) and genuine homogeneity at the realized property level (the level of mental state kinds). From a conceptual standpoint, they are absolutely right. Multiple realization is more or less the thesis that some type of thing is multiply realized in various types of other things. With respect to the issue about the relationship of the mind to the brain, in particular, multiple realization demands that mental state types are multiply realized by brain state types (or perhaps other types of things—states of some machine, or states of some Martian's 'brains'). Clearly then very much needs to be said about how one classifies mental states and brain states. Without offering some independent way of slicing up mental state types and brain state types, the argument against identity theory will merely have a hypothetical feel. One would like to say, without such work in place, that, *if* multiple realization were true, that would of course rule out identity theory—but that we really have not quite established that multiple realization is really true. That very little literature exists about how we should do this classifying of the relevant states is not only a bit confusing, but is also, I hope, a feature of this dissertation that is relatively novel (if not helpful).[3]

A related conceptual issue here, not explicitly touched on by any of the criticism I discussed above, is the realization relationship itself. Shapiro explicitly worried about the *multiple* half of multiple realization, but, of the works I have referenced so far, he does not say anything about the other half: the nature of realization itself. In the same way that it is surprising that few writers have had much to say about what mental state types there are, or what brain state types there are, very little literature exists on the nature of realization. For my part, I plan to avoid the issue as well. There is enough to be getting on with not to have to tackle that problem,

[3] To be as clear as possible here: I will not, by any means, want to offer a *general* account of the nature of types or kinds. I only want to ram home the point that we must appeal to the work in different sciences in order to establish how they classify the states I am concerned with.

too. Instead I will focus much more on the issue about how to distinguish multiple realizations. Since I do not think that much of what I want to do hangs articulating a view on realization, I hope the reader does not mind. My hope is rather to argue for a particular classification of mental and brain state types, then show that empirical work has been done which shows that the mental state types stand in a one-to-many relationship to the brain state types. I think it *will* be necessary to engage Shapiro's worry about what constitute *multiple* realizations—and I plan to do so. But I am willing to adopt any theory of realization which can accommodate the conceptual and empirical work I advocate and use. I do not think that I am obliged to make any commitments. There are of course, as everywhere in philosophy, many issues tangled together so that commitments here spell trouble over there. In particular, with views on realization, there are worries about mental causation and individualism (about the bearers of mental states). So there might be some concern that the development I give to the multiple realization argument will inadvertently pin me to a strange metaphysics. If this turns out to be so, I will accept whatever metaphysical consequences my position requires.

That leaves the empirical side of things to explore. In light of Bechtel's and Mundale's work particularly, it would be nice if I could adduce some examples from neuroscientific journals which support the basic idea that mental state types stand in a one-to-many relationship to brain state types. Finding some empirical results across species, or within an individual, which support the existence of this relationship would go a long way towards responding to their concerns. As I have already hinted at, this is certainly something I hope to accomplish below. I believe the conceptual work of getting clear on the relevant types must come first, though. How else could we objectively interpret any empirical findings if we did not come to agreement on the types to begin with? But, after doing so, I will try to show how certain cases fit the conceptual parameters discussed (i.e., given how we decided to type the relevant states, and given the one-to-many relation required by multiple realization, I will present a few cases which do meet those conditions).

Hopefully, too, this will allow me to form a general critique of the line Bechtel and Mundale take. That is, if I can empirically defend the idea that mental state types *are* multiply realized, then something must be wrong with their suggestion that successful neuroscientific practice has depended on the truth of the identity theory. And it would be nice if I could explain what has gone wrong. So the middle part of the dissertation—after having worried about the classification of the relevant state types—I would like to do just that: explain the ramifications of my view in the context of the prominent critics. In the same way that I would like to spend a few pages detailing what I think is wrong about the line taken by Bechtel and Mundale, I would also like to explain, for instance, why Polger's attack on Putnam's intuition is also no good. Again, the main reason I decided to attempt this project is my dissatisfaction with the existing criticism of Putnam's and Fodor's old argument. Everything I read always left me completely unmoved. Taking up a good chunk of the dissertation to justify that dissatisfaction by meeting the critics is—to say nothing of any academic obligation—precisely what I want to do.

Chapter 2: What is a Mental State Type

I want to start first with the work of clarifying mental state types and brain state types. It may seem like this is easy business, but looking at a particular empirical case might help show how difficult these notions can be. This case has turned out to be very central to debates on multiple realization and I will discuss it extensively throughout this dissertation, particularly in the last chapter. But its value here is to show how a seemingly obvious case of multiple realization is held hostage to a treatment of mental state types and brain state types. I hope it thus helps to motivate the importance of the present chapter and the next.

Sur et al (1998) rewired a few thalamocortical connections in ferrets in order to make different cortical areas serve different processing tasks—e.g., they rerouted information along the lateral geniculate nucleus (originally destined for the visual areas of the brain) to the primary auditory cortex, actually achieving a limited kind of vision in the ferrets. One might be inclined to say this case is an obvious example of multiple realization. If a different area of the cortex (than normal) is processing certain information in order to establish a form of vision, it might seem like the same kind of mental thing is being processed by a different brain kind. Shapiro (2004) and Polger (2009), however, disagree that the rewired animals count as evidence for multiple realization. They dispute both that the vision achieved in the rewired animals is of the same kind *and* that the new areas of the cortex (into which the sensory information was rerouted) should count as distinct brain kinds. As it turns out, the ferrets only achieved a very limited kind of vision and the newly rewired auditory cortex became organizationally more similar to the primary visual cortex. Hence there were some grounds for a reevaluation of the status of these ferrets as evidence for multiple realization—all of which turned precisely on when the mental side of things was genuinely the same and when the brain side of things was genuinely different. Clearly there is a reason for pausing over just what makes two mental state kinds the same, or two brain state kinds distinct. Without coming to some sort of independent understanding of how to individuate these states, it will be difficult to settle on just what the ferrets have to say to us vis-à-vis multiple realization.

In order to gain more traction here, I will start with the mental state side of things. Now the move I want to make immediately, to get to the heart of the issue right away, is to forget about 'mental' states altogether and replace that kind of talk with *psychological* states. Of course I am aware that framing the discussion in terms of psychological states rather than mental states somewhat shifts the scope of the present work away from the classical concerns of Place, Smart, and original identity theorists, but I think this switch provides the best way of profitably evaluating the multiple realization argument. The rest of this chapter is a sustained answer, looking at the matter from different perspectives, to the question of why the switch is a good idea.

Section 2.1: Why Psychological State Types Instead of Mental State Types?

First of all, talking about psychological types instead of mental state types appears to be an accepted move in the rest of the literature. On the *very first page* of Shapiro (2008), Polger (2009), and Bechtel and Mundale (1999), 'mental' and 'psychological' are both used to formulate the identity theory and/or the multiple realization thesis. Here is a sample of some textual evidence that shows these writers do not shy away from 'psychological' kinds (at least in the papers cited). Polger's contains this rendering of the identity theory (pg. 458): "In particular, the mind-brain identity theory is a theory of kinds or types: it says that psychological kinds are identical to neuroscientific kinds". Shapiro's essay begins with the question (pg. 514): "Are psychological properties multiple realizable?" So, to some extent, my desire to lay out the whole issue in terms of psychological states should not be controversial—at least not with respect to how philosophers who care about these things usually talk. There is, I suppose, the possibility that some of these writers (though probably not all—like Bechtel and Mundale) are using 'psychological' in a loose enough sense that it can be taken to mean 'mental'. When one considers that the examples of psychological types they go on to mention (like 'pain'), it is not obvious that they mean psychological in any *scientific* sense like I do. But, at the same time, as is clear in Polger's quote, these writers also invoke neuroscience, which, I take it, probably cannot be employed in any non-scientific sense. If that is so, then there is reason to think they are framing the MR debate in terms of the relationship between the posits of two scientific disciplines. Additionally, there are many cases in the literature where it is quite plain that 'psychological' is intended in the scientific sense. The already mentioned case of the neonate ferrets is a good example: the *lack* of psychological similarity between the ferrets is unquestionably determined from a perceptual psychological standpoint.

But then does it really matter which way of talking I prefer? I think it does at least a little, and I think so for a few reasons (beyond the fact that everyone already does employ talk about psychological kinds). So at the risk of belaboring a bit of a tempest in a teapot, I want to say something about those reasons. One thing that talk about the mental immediately connotes is phenomenal consciousness. That is, when we phrase things in terms of mental states, there is a strong temptation to view the individuation of the relevant states as accomplished by appeal to qualia, or the felt aspect of sensations. Take Polger (2002) as an example. He is one of the many who tries to combat multiple realization by hammering home the point that different creatures may not feel the exact same pain that we do. To remind the reader, the strategy here is to show that two individuals with admittedly different types of brains do not actually share the same mental state types, thus nixing any possibility of one of the former being multiply realized in many of the latter. Polger recognizes that we could have a strict or a fairly loose understanding of sameness of mental state types. On the loose reading, we would just be claiming that other creatures share the same general kinds of (phenomenally) conscious states that we do, while on the more strict understanding we would say that these creatures have the exact same kinds of conscious states that we do.

He uses the word "empathetic" to get at this notion of 'exact same kind'. If a dog, say, were able to share the exact mental state we have when we are in pain, then we may say that we could empathize with the animal's pain (in virtue of sharing the same kind of mental state). On the other hand, if the dog, while having some sensation or other while in 'pain', did not feel the same way that we did, then we would say its mental state was of a different kind. When we remove the nonhuman species in question even farther from us—say, to birds or octopi—it becomes prima facie doubtful that these species empathetically share our same mental state types. This is not necessarily the same thing as saying such creatures do not feel certain sensations and have certain experiences. But to suppose that they feel the exact same kinds of sensations and possess the same qualia that we do intuitively seems implausible. Maybe we do not even need to compare across species, in fact. It does not seem crazy to suppose that, from human to human, there are actually some differences in exactly what it feels like to be in pain. When we think about other types of sensations and experiences, the case may even seem compelling: for instance, to use Polger's example, we do seem to believe that expert musicians hear something different from the novice while listening to music. Maybe this point is enough to make reasonable the claim that there is some intrahuman variation of felt experience. If that is so, then an empathetic understanding of multiple realization of mental states cannot even get off the ground.

Now, what the schema for individuating mental state types starts to look like is something like the following: two mental states, X and Y, count as belonging to the same type if and only if undergoing X is exactly (i.e., empathetically) like undergoing Y. And depending on how strict or loose we want to be, we will strengthen or relax the notion of 'is exactly like.' I am not claiming that this is precisely Polger's view on mental state type individuation, nor am I trying to pin it on anyone else. It is just a handy way of illustrating an intuitive picture of how to isolate the realized state types when the focus is on qualia. But the obvious problem with this view of how to single out the mental state types is that not all mental state types *could* even be typed by qualitative, phenomenal features. Beliefs, for instance, do not appear to have any 'feel' to them. So in a sense undergoing one belief would be exactly like undergoing another, meaning we would have to type them as the same kind of state. This is an unacceptable consequence. One is obviously free to keep on talking about 'mental' state types, but the focus will have to be on other criteria for individuation than just the phenomenal aspect to our mental states. How do we then go on to articulate what those criteria should be? What is our next schema going to look like for identifying these mental state types?

I cannot see how the answer to these questions will avoid an appeal to psychological theorizing. There are, of course, many psychological models which posit certain states and mechanisms as causally responsible for ours and other species' gross behavior and cognitive performance. Sameness of mental state type would thus amount roughly to our ability to subsume two individuals under the same psychological theories (vis-à-vis some behavior or capability). That kind of judgment of sameness, though not necessarily as precise as possible in every case, would be easy to establish and would not so obviously run the risk of poorly individuating all of the relevant state types. This seems like a decent reason for why the literature equivocates between talk of psychological types and mental state types. Writers sometimes talk about mental state types, but when they do ever bother to explain what matters for their identification, immediately the criteria for individuation that emerge are psychological. Again one needs to look no farther than any philosophers' commentary on the neonate ferrets. Particularly the criticism of these ferrets as evidence for multiple realization usually revolves around the psychology of vision and how the ferrets' abilities are not (psychologically) comparable. But if that is the way to resolve the problem about how to type mental states—by looking at how the science of the mind picks out its kinds—then there is no obvious reason why we should not just call them 'psychological types' instead of the more ambiguous, possibly misleading 'mental state types'.

But there is another main reason, besides the possible troubles with mental state type individuation and the apparent fact that the relevant philosophers of mind seem to already have done so, for why we should swap out 'mental state' for 'psychological state': the very history of multiple realization is intertwined with the 'reduction' of psychology to neuroscience or physics (not 'mentality' to 'brains'). Unfortunately, reduction means a lot of different things to different writers; it is one of those annoying terms not used univocally in philosophy. There are, to give some examples, a couple of different common senses of reduction: methodological reduction, theoretical reduction, and explanatory reduction. Now, it is not too important for my purposes to distinguish all of these notions of reduction, but at least explaining one of them will give the reader a taste for the kind of issues that are raised in this context. Hopefully that will also give the reader a sense of why it would be relevant to construe the main argument as involving *psychological* state types as multiply realized in brain state types.

Let us arbitrarily take the theoretical reduction to be our example of reduction. If it helps the reader, this is the notion of reduction I take to be Fodor's target in his (1974). This sense is motivated by the tempting idea that physics must be general with respect to all of the sciences. That is, if everything is physical at bottom, then we should expect, it seems, the laws of physics to cover all of the phenomena we observe. And given this generality expectation, we should thus find there are so-called bridge laws, composed of the vocabulary of the special science and physics, which deductively connect the laws of the reduced science to the laws of physics. This deductive relationship from the reduced theory to the reducing theory is what is responsible for the 'reduction' of the former to the latter. One might also expect that same type of reduction—in terms of the laws of the sciences—to hold between psychology and neuroscience. Multiple realization is invoked here in order to show that the kinds picked out by the vocabulary of the special sciences do not stand in one-to-one relations to the kinds picked out by the physical vocabulary. In this way, the 'bridge laws', which are necessary in order to reduce the special science laws to physical laws, are not forthcoming. In particular, since it is taken that laws are composed of the *kind* predicates of a science, and since the correlation between special science kinds and physical kinds is one-to-many, and finally since disjunctions of physical kinds typically do not (if ever) compose yet another physical *kind* (recall Fodor's distinction between kinds and properties), there cannot be bridge laws in order to provide the reduction. The multiple realization of the special science kinds blocks the possibility of formulating them.

What Fodor chiefly hopes to gain from this move is a way of reconciling a commitment to physicalism with a denial of the generality of physics vis-à-vis the special sciences. It had seemed to many that a commitment to physicalism—which, of course, was non-negotiable—was just another way of saying that really every scientific kind was coextensive with some physical kind, that the special sciences were really quicker, less precise ways of doing physics. The multiple realization objection to such a reduction is a way of showing the indispensability of the special sciences. Since the world is organized in such a way that the same regularities are hit upon by heterogeneous physical combinations, and since we should try to state the generalities that there are to state, we have strong reason to reject the generality of physics. But of course there is no dualism here either: each token of a special science kind is identical to some token physical kind. So the commitment to physicalism is not jeopardized.

Admittedly it is not my direct goal to weigh in on this debate about theoretical reduction. But it is easy, in light of the foregoing, to see why anyone might be interested in the potential multiple realization of psychological state types with respect to, say, brain science. If it does turn out that psychological state types are multiply realized in brain state types, then not only can we close the books on identity theory, but we will also have an interesting result about how the theoretical postulates of the sciences of the mind relate to those of the sciences of the brain.

Since there does exist a historical precedent of worrying about the multiple realization of psychological states (about which the conclusions I draw here would certainly be relevant), and since when push comes to shove there seems to be a reliance on psychology to resolve sameness of mental state types, I conclude that a shift from individuating mental state types to psychological state types is well-motivated. This is without even mentioning the really important bonus of typing states by looking at the sciences: that it is possible to be quite precise about when two tokens belong to the same psychological type. I discuss this reason for adopting a psychological notion of the mental state types below in section 2.3.

Section 2.2: But What do Psychological State Types Look Like?

Then let us take for granted this amendment to the multiple realization argument. That still leaves the obvious question: what are the psychological state types? Establishing mental state continuity between two individuals seems hard, but how will things work in establishing their *psychological* sameness? Well, it strikes me as the obvious strategy simply to take a look at what psychological models of various processes are saying. In particular, by looking at the operations and mechanisms posited by the best going psychological theories, we should be able, first of all, to regard the psychological sameness of two individuals as a function of their behavior being predictable and explainable by that psychological theory. Secondly, in understanding the theory in question, we should be able to figure out the relevant state types quite straightforwardly. Part of what theories do is to specify the relationships that hold between certain types of states of affairs in the world. Let me pause for a moment and say more about each point in turn.

The first point about establishing psychological sameness is simply taking a page out of the book of scientific realists. Scientific realism is basically a two-part view. The first part, an epistemic part, is that we are justified in believing in the postulates of an acceptable scientific theory or hypothesis. The second part, the metaphysical part, holds that the phenomena contained in scientific theories, the real meat and drink of the theory, are 'real' and do have the properties ascribed to them by the theory—and not just from a scientific viewpoint; the phenomena are ‘out there’ regardless of our pragmatic concerns. The metaphysical point about the acceptability of a theory providing a reason to posit the existence of certain states is clearly useful for my purposes. Now I only have to say that two individuals instantiate all the same (psychological) state types by showing that the same (psychological) theory applies to them. If the laws of the theory subsume the behavior of each individual, scientific realism guarantees that the individuals really have all the same states and kinds mentioned in the theory.

We all know, of course, that there is a substantial line of criticism of scientific realism (see Van Fraassen, 1980). In particular, there is the problem of two theories being empirically adequate—accounting for all the observable phenomena—and yet remaining so even while making appeals to different sets of unobservable phenomena (different theoretical entities). Hence, so the argument goes, there is no reason to prefer one theory’s unobservables to the other; the observational evidence simply underdetermines the choice between the different theories (vis-à-vis the theoretical entities). Given the scope of my dissertation, however, I am going to step over this worry. From the literature it seems like the standard view is against those like Van Fraassen. So I will take that as a strong enough reason comfortably to overlook the critical literature and just assume scientific realism.

To understand what I mean about the second point about what the psychological kinds are, compare Fodor’s remarks (in his 1968) about the notion of ‘behavior’ in psychology. Prima facie, there may appear to be a bit of a problem. The average meaning we have for ‘behavior’ is something rather like ‘deportment’. But it is clear that dreaming and problem-solving, say, count as behaviors for the psychologist, though neither have anything to do with behavior construed as ‘deportment’. So, since psychology is at least *partly* the study the study of behavior, someone had better say something informative about what counts as behavior in the psychological sense.

When you consider, say, a body falling from a cliff is also a kind of behavior—just as much as the behavior of perceiving the edge of a cliff, which seems patently more psychologically interesting—the question becomes even more pressing: just what constitutes the domain of psychology? Or, to put the question in different words: what is it about two events that make the one relevant, the other irrelevant for the purposes of psychologists? The response is to say that all sciences eventually define their respective domains as they go along. Or, as Fodor puts it: "there is . . . an important sense in which a science has to discover what it is about" (Fodor, 1968, pg. 10). It is only through the articulation of laws and the making of predictions which allow us to see what important similarities obtain between what pretheoretically seem like heterogeneous events. That is how you figure out what counts as the 'behavior' which defines the kinds of events for which psychological theories are required to have an explanation.

So, too, I want to say with identifying the *kinds* of psychology: if you want to know what the important psychological types are, then look at the laws and predictive generalizations derived from the theory. Those laws and generalizations are what explain the important relations which hold between the theoretical terms also postulated by the theory (which pick out the importantly similar events in the world). Those theoretical terms also pick out the psychological kinds. So, for instance, if there were some true law relating 'drives' with 'motives', we could identify those things as psychological kinds. However it is that laws get confirmed, assuming that law between drives and motives *is* confirmed, it describes an interesting connection between types of events in the world. Those events are classed together by the science to name its particular kinds. Hence, the problem of finding out what psychological state types there are reduces to the understanding of accepted psychological theories.

Of course looking at accepted theories in psychology is not, at least in principle, cause for concern. This is one of the reasons, as I mentioned above, that I prefer the move away from classifying mental states to classifying psychological states. If two individuals' behavior and cognitive performance are explainable from a particular psychological theory, that is evidence for believing that the same laws and generalizations hold for those individuals, which is in turn evidence for believing that they instantiate the same psychological state kinds. All we need to do is see what the psychologists are telling us and we can close the book on the issue of the sameness of the multiply realized state kinds.

Section 2.3: The Pros and Cons of Computationalism

But there is a bit more we could say about current psychological theories in general, which will also bear on our evaluation of sameness of psychological state type. In particular, whether you do neuroscience, cognitive neuroscience, cognitive psychology, or do any research remotely related to the brain, there is a general consensus that the brain is, in some loose sense at least, an *information processor*. Through the sensory organs, the brain is sensitive to information about the distal world carried by many different media: for instance, electromagnetic radiation, oscillations of pressure transmitted through the air, and the chemical composition of substances (for tasting and smelling). After this information is transduced to electrical-chemical signals (by, say, the retina or the cochlea), the brain itself exploits the information carried by the distal stimuli, making that information available for the organism to perceive, form beliefs, and ultimately make the decisions for actions which keep it alive. Given the also near universal commitment to the idea that psychological capacities are housed in the brain, it seems like a decent bet that respectable psychological theories are going to be in the business of shedding light on the nature of the information processing in which the brain is engaged. And this is what we find in contemporary psychological theorizing. The theories one finds are concerned with postulating state types within processes that correspond to the kind of information which the brain manipulates and how exactly it manipulates that information. To use some vocabulary with which philosophers of mind will be familiar, current psychological theories assume a *computational* view of the mind, a view that says cognitive processes are computations performed over representations that carry the all-important information. Virtually every psychological model one finds is neck-deep with talk about representations. Unfortunately, however, invoking computation in this context has positive and negative implications. I will begin with the negative implication, arguing briefly that it should not actually worry us, and then go on to the virtues.

This negative point of invoking computationalism derives from a point made by Polger in his (2002). In evaluating the multiple realization argument, he insists that we cannot presume a particular theory of mind in order to justify our intuitions about the kind of multiple realization we find plausible (e.g., SETI-MR, or Standard-MR), otherwise the multiple realization argument simply amounts to the statement of the plausibility of that theory of mind. More specifically, he says, ". . . if the plausibility of multiple realizability depends upon the plausibility of functionalism, then the argument from multiple realizability only repeats the assertion that functionalism is plausible" (Polger, 2002, pg. 147). In other words, we would like a particular kind of multiple realization claim to be true, say Standard-MR. But if the only reason I can cite for the plausibility of Standard-MR is my belief in the plausibility of functionalism, then I have not done a whole lot to advance the issue about which theory of mind (not) to adopt. The multiple realization argument is supposed to settle any questions about the truth of identity theory, not presume its falsity from the beginning. Sneaking in the assumption of functionalism's truth trivializes the strength of the multiple realization argument by counting merely as a reassertion of the plausibility of functionalism.

The problem for my view, then, is that replacing mental state types with psychological state types, given the nature of current psychological theories, compels me to adopt a computational theory of mind. But computational theories of mind are just a species of functionalist theories of mind—where the important, type-individuating causal relations are all computational. So now it appears that I am open to Polger's question-begging charge of sneaking in a functionalist theory of mind in order to motivate the truth of multiple realization. Although I will come back to this problem later in the section 6.1, where I consider it in more detail, I want to canvass a few short responses to tide the reader over for now.

First of all, the criticism he is advancing here is just not compelling. Functionalism surely *does* constitute—at least prima facie—a plausible theory of mind. Yes, the view also seems to allow for the possibility that systems of indefinite physical composition could share our mental state types. And if turns out that such an implication is incompatible with another possible theory of mind, then we should be interested in confirming that kind of incompatibility. But in this context it does not seem like an *illicit* move to assert the plausibility of functionalism. As long as functionalism—particularly in its computational form—has independent empirical plausibility, it is not clear that a question-begging charge carries much weight at this point in the dialectic. It is this line which I will unpack in much more detail in 6.1.

Also, asserting functionalism does not *rule out* the truth of the identity theory—i.e., asserting functionalism does not amount to asserting the falsity of reductive materialism. It is compatible with a functionalist theory of mind that there simply is only one realizer for whatever functional role is specified—just as they say sometimes that there is only one man for the job. Kim's views on 'functional reduction' (see his 2005) seem to be just such a compromise between functionalist and reductive theories of mind. Lewis' position in his (1972) is another example. Even if computationalism allows for multiple realiz*ability*, it does not yet show that psychological kinds are in fact multiply realized. If that is so, then it is not obvious to me what is wrong with having a functionalist position at the bottom of our intuitions about how likely we think multiple realization is. Advocates of the multiple realization argument are still on the hook for arguing for the truth of the premise from the argument from the introductory chapter (pg. 11). The evidence I want to provide for psychological state types being multiply realized is not that a functionalist theory of mind is compatible with such a state of affairs; that does not make any difference to me at all. The question of the truth of multiple realization is, after all, an *empirical* one—and I will have to provide empirical evidence that psychological state types are multiple realized.

But the strongest way of replying is to point out that I am not using a computational picture of psychology to show anything at all about the identity theory (that, e.g., it is false). I am only adopting (a computational notion of) psychology in order to better pin down the nature of the higher order state types which are realized in brain state types. And I can do this without making any under-the-table metaphysical bets. Later I do use phrases like 'algorithm', 'strong equivalence', and 'information processing.' These words appear to be metaphysically loaded against the identity theory, but I do not require any of that metaphysical baggage from these terms in order to put them to polemical work. I only want to use these phrases to know better why two token psychological states are of the same kind. If you know that two mental processes are, say, ‘strongly equivalent’, then you know that the representations involved at a given step in each process are of the same kind. Or, somewhat less complicatedly, the simple act of specifying the ‘effective procedure’ for carrying out a psychological task is going to invoke representations at every bend and transformation of that procedure which immediately pick out psychological kinds. To that extent, I can even claim agnosticism about the relationship between computational psychology and functionalism/identity theory. Maybe the computationalists are functionalists, or maybe they think there is something really unique about the way in which a brain computes our mental processes. It does not matter to me either way. If computationalism implies multiple realizability, I still have to provide evidence for a psychological kind really being multiply realiz*ed*—and nothing about my adoption of a computational picture of psychology (for the purposes of figuring out which state types are the same) is going to get me there on its own. I do not need the concepts to settle any metaphysical disputes; I need them to help individuate psychological kinds at a fine level of grain. From that perspective, in fact, the adoption of a computational psychology actually does burden me with a substantially strict notion of sameness of kind. For now I have to show not only that two distinct creatures can see, but more specifically that the processes which they use to achieve their vision are at least roughly equivalent. This is, I think, a charitable *concession* to the identity theorist—not a question-begging piece of theft.

But this is also the main *positive* feature to take from current psychological theories being computational in nature: that we can tighten down further the notion of sameness of psychological state types. In particular, we can make sense of the demand for the *strong equivalence* of the computations in question (see Pylyshyn, 1984). This idea of strong equivalence is that two processes count as the same when they use, as it were, the same algorithm for processing the information they do. For instance, one could do addition by various different algorithms. There is the old-fashioned and quick method of reading the numbers from right to left, where numbers over 10 are carried over to the next column, revealing the sum from right to left. But there is also the 'partial sums' method where the adding is actually done from left to right. The numbers in the left most column are added together with zeros for all the numbers to the right, then that number is added to the sum arrived at by the adding of the column immediately to the right of the first column, and so on. The computations will churn out the same ultimate sum, but they do so according to different, though still effective, methods. In other words, the computations are at least *weakly* equivalent: they take the same input (the numbers to be added) and produce the same output (the sum). But these different ways of adding do not get from input to output in the same way; they employ different procedures for getting the job done. Two *strongly* equivalent procedures, on the other hand, get from input to output using the exact same method. The point worth taking, for our purposes, is that we should require the equivalence of our two individuals' psychological processes to be strong. It would not be enough for genuine sameness of psychology if two organisms were capable of arriving at the same outputs (given the same inputs), but did so in algorithmically different ways.

For a quick example, it is well known that in humans there are two 'streams' in visual processing (Ungerleider and Mishkin, 1982): a dorsal stream and a ventral stream. The dorsal stream is responsible roughly for processing information about *where* an object in our visual field is. The emphasis is not so much on the identity of the thing that is moving around, but rather with keeping track of the location of it relative to the location of body. On the other hand, the ventral stream is roughly responsible for processing information about the identity of the objects in our visual field. Contrasting with the dorsal stream, there is little emphasis on the movement of the object; the point is to figure out *what* is in the visual field. The information is later integrated to give us the downstream, unified experience we have (of, say, a Maserati—the identified object—driving down a street—the movement of the object), giving us a rough sense of the manner in which the human brain processes information from the retina. Intuitively, it seems that for a creature to share our visual psychological states they would have to process the information we do—i.e., use the same algorithm we do for visually figuring out what is in our environment. Part of this equivalence of the processing would involve breaking down the flow of information into two different streams of analysis.

Another way of putting the point is that a creature with exactly our visual capabilities would, among other things, have to be susceptible to the same illusions and agnosias which afflict humans. Prosopagnosia, to pick an example, is roughly the impaired ability to recognize faces. There is no problem with locating where faces are (when they are moving, e.g.), but it becomes hard to identify them. If a creature is unable to have a visual disability like this, then we have some evidence that their visual processing is not strongly equivalent to our own. That is, you can only get disabilities of identifying objects while having no problems tracking their movement if the processing of the respective information is separated in some way. So in order to process the information humans do, in the exact way they do, it is at least necessary to separate the processes of identifying objects with tracking their movements. This is a simple example; the nature of the processing of visual information is hugely complex. But it gets us closer to understanding what is meant by strong equivalence of the computational procedures that underlie our psychological capabilities.

So, again, this feature of contemporary psychological theories—their being computational in nature—helps to establish a stronger account of the sameness of the psychological state types of individuals. A possible worry, at least with a more intuitive way of typing (conscious) *mental* state types, is that the notion of sameness is arbitrary or imprecise. We do not doubt of course that other people or animals are in pain, for instance. But it certainly could be doubted that the way the pain feels is similar across different species or even different members of the same species, so that two different individuals in pain could not quite be said to instantiate the same mental state type. The virtue of computational psychological theories is that we *can* nail down a much more precise notion of sameness of state types. If, say, the entire repertoire of the visual abilities (and disabilities) of two different creatures are explainable and predictable via a particular theory of vision, then we have excellent grounds for maintaining that those creatures share the same visual states. More particularly, we have good grounds for maintaining that the exact nature of the processing of the information, the computations performed over particular representations, the state type to state type transitions required by the algorithm, are the same. That the theory can explain what visual stimuli the creatures are both capable of distinguishing, that it can predict what sort of illusions the creatures are both susceptible to, that it can predict how a particular interruption of the processing produces the same particular deficits in both creatures, counts as the best reason we could have for asserting the sameness of the psychological state types of the creatures in question.

<u>Section 2.4: An Example of a Psychological Process</u>

So let me take the time, then, to provide a somewhat extensive, concrete example of a psychological theory of some ability. This will provide an illustration for the reader of some of the themes I have been discussing so far: viz., how strong equivalence of the computational processes falls out of a theory, how precise notions of psychological state types (kinds) fall out of a theory, and how employing a theory to explain particular phenomena generally helps to show that two creatures instantiate the same processes and state types.

Picking almost at random, I will look at a theory about how we recognize spoken words. Note, in order to make things as clear as possible up front, that we are not considering how we recognize *written* words. These processes are, according to popular theories, quite different, and it makes sense why they should be. That is not to say there are no similarities between the two. For instance, whether we are recognizing written or spoken words, the stimuli for each can vary widely. We understand the same words written in endless kinds of font, and we understand words spoken in different dialects, different speeds of speech, and even when the speaker has a cold (i.e., is speaking differently than normal). An important difference, however, is that the written stimulus is there quite a lot longer (indefinitely almost) than the acoustic stimulus, which is present for a flash. Also, the perceptual apparatuses are quite different: you need, e.g., a retina to recognize written words and a cochlea to recognize spoken words. Note also that the example I have chosen is quite specific. I am not concerned with our capacities for language in general, our abilities to produce language, or even our abilities to comprehend language. I am only talking about a very specific ability: to recognize a single spoken word.

The general picture of the theory in place divides the processes involved into three stages: an *initial contact phase*, a *lexical selection phase*, and finally *word recognition*. There is much to explain, particularly to do with definitions, but I will start with a general outline of the phases, then move into the details from which we can start to understand the psychological kinds to which we are committed in accepting this picture. The *initial contact phase* involves the encoding of purely acoustic information for lexical selection. That is, this part of the process is involved in, so to say, interpreting the sound waves stimulating the perceptual organs of the inner ear into the phonetic representations out of which lexical selections are made. The *lexical selection phase* is the part of the process devoted to matching the incoming pre-lexical, sensory representations with some representation in the lexicon. At the endpoint of this 'matching' phase, we have *word recognition*, where the foregoing processes all accumulate in picking out a single entry in the lexicon.

To start nailing down some of the details of this view of spoken word recognition, let us start with the initial contact phase. A key concept here is that of a 'phoneme', since the typical understanding of this part of the process is that we transduce the acoustic information present in the sound waves into representations of phonemes. Essentially, a phoneme is the basic unit of sound *in some language*. For instance, /r/ in '*r*un', or the /l/ in '*l*augh' are phonemes in English. To give some more interesting examples, take 'ju*dg*e' and 'reli*gi*on', both of which contain the phoneme /dž/. In order to understand more precisely how these sounds are supposed to constitute the basic units of sound in English, compare them with *phones*. Phones are kinds of sounds considered outside of contexts of language. The [l][4] in 'laugh' and [r] in 'run', while obviously corresponding to distinct sounds (independent of some lingual context), are actually the same phoneme in Japanese (i.e., considered in the context of that language)—hence the notorious pronunciation difficulties for native Japanese speakers using English. So, to be precise, the initial contact phase is supposed to be about translating acoustic properties carried by sound waves into representations of these phonemes—that is, not necessarily faithfully recording the language independent sound present, but rather capturing the phoneme which best matches the sensed stimuli.

[4] Customarily, phonemes are indicated by slashes (e.g., '/x/') and phones by brackets ('[x]').

Once the phonetic representations have been picked out, we are ready to compare that information to representations stored in the lexicon. This is the part of the process known as lexical selection. To understand this portion a bit better, we need to know in the first place what the lexicon is. In brief, it is like a mental dictionary. The general psychological understanding of this concept is that the lexicon contains all the linguistic information we know about a word: its phonology (what it can sound like), its orthography (what it can look like written), its semantics (what the word means), and finally all the syntactic information (the grammatical facts about the word). So, when we say the phonetic representations are compared with representations in the lexicon, we mean to say the encoded information about the acoustic stimuli is compared with our mental dictionary—for example, the parts associated with phonological information we have about words.

The received view on how this particular process goes is the 'cohort model' (see Marslen-Wilson, 1987). The rough idea is that we set up a 'cohort' of possible words which could be picked out by the train of phonetic representations, eliminating words in the lexicon which do not match the incoming acoustic information until only one word is remaining. More precisely, the beginning of a word sets up what is called the 'word-initial cohort'. This cohort will be presumably quite large if, say, the initial phoneme represented is /r/ (i.e., there are a lot of words that being with that sound). But it will be whittled down quite quickly on the basis of more phonological information. Now, depending on how autonomous the processing is—how, to use Fodor's (1983) language, *modular* the processing is at the earlier pre-lexical stages—we might also include syntactic and semantic information contained in the context as relevant factors for eliminating members of the cohort. For instance, if what should appear in this particular part of the spoken sentence we are hearing is a noun, if we have been talking about sewers, and if the next phoneme is something like /æ/, the word 'rat' is probably going to be the last word standing (even before we have heard the whole stimulus). This employment of syntactic and semantic information is supposed to constitute a top-down effect, where higher-level processes (processes integrating lexical selections with their learned semantic and syntactic roles) affect lower-level processes (the bare transference of acoustic information into the phonetic code for finding a match in the lexicon). The existence of such an 'interactive,' non-modular network of processes is still a contentious subject in this area of psychology. For our purposes, though, nothing much hangs on how we want to construe the relevance of contextual, higher-level processes. Remember, the main goal here is simply to provide an example of a psychological theory of some capacity or other (in order to get a grasp on how to type psychological kinds and what constitutes strong equivalence of two algorithms). We do not yet have to make sure the scientific details are all worked out.

After the cohort has been reduced to the distinctive entry in the lexicon, we have reached what is called the 'uniqueness point,' the point at which a word can be distinguished from all similar words. Depending on your view about the relative autonomy of the (pre-lexical) processing, the uniqueness point may not correspond to word *recognition*. If semantic and syntactic contextual factors make a difference in, e.g., the lexical selection process, then one might be able to recognize a word before enough phonetic information is present uniquely to identify some entry in the lexicon. If we think the evidence favors a more modular view of the early processing stages, then word recognition co-occurs with the uniqueness point. From here a new set of (lexical) representations handle the integration of the word selected with higher-level processing (having to do with employing the word in whole sentences—e.g., using the word in syntactically correct ways, or placing the lexical representation in a complex representation of the whole sentence).

So that is how the psychological story is supposed to go for spoken word recognition. There is, roughly, a pre-lexical code used to get the acoustic information into a lexically appropriate code and then an examination of the lexicon in order to pick out the right entry (corresponding to the spoken word), and a post-lexical code used for integrating the lexical selection with other higher level information. That is a standard psychological theory for accounting for our ability to perceive speech.

Now what precisely was the point of talking at length about word recognition? There are a few reasons. First of all, we now have a much clearer picture of the nature of the processing involved in how humans recognize spoken words. More particularly, we have a much more accurate picture of the actual *algorithm* used to recognize a word. We know, for spoken word recognition, that the input is acoustic information and the output is recognition of a word, but we did not know precisely how the brain managed to get from one to the other. The above model is explicit about just how the acoustic information is exploited to generate word recognition. Of course, we needed to understand the algorithm utilized so that we could make sense of strong equivalence. A computer (or organism, or whatever) might recognize some word, say, not through a pre-lexical representation of the phonemes instantiated in the acoustic signal, but rather through some direct analysis of the acoustic information to words. Alternatively, the computer (or organism, or whatever) may employ phonetic representations as intermediaries between the purely acoustic information and post-lexical processing, but does not use the cohort model of lexical selection in order to pick out the right lexical entries. Such processes—though equivalent at least in a gross input/output, sound-to-word way (similar to how the methods of addition start with the same addends and arrive at the same sum)—would not be *strongly equivalent* to the exact processes involved in human word recognition. Second, looking at this (or any) theory helps us to type psychological kinds. In this case, we have a few different kinds: pre-lexical representations, post-lexical representations, the lexicon, etc. Of course, from a computationalist perspective, the thing that counts for individuating states are the representations involved in the processing and how they are transformed from computation to computation. In the case of spoken word recognition, the going theory posits a dual code—post- and pre-lexical. Hence, the important (computational) psychological states involved are, roughly, phonetic representations (used to identify distinct lexical entries via the cohort model) and lexical representations (which, when selected, are integrated, e.g., into larger grammatical constructs).

To tie this all together, remember the overall project: we are trying to pick out the psychological state types *and* show how two individuals can share them. I have been suggesting that we use the best psychological theories in order to do the work for us. We can, that is, evaluate sameness of two individuals vis-à-vis psychological state types by looking at what the psychologists are saying about a particular psychological ability. With a theory in hand, all we have to do is see whether the predictions and generalizations made by the theory subsume the behavior of different creatures. If they do, assuming scientific realism, then we have reason for thinking that those creatures really do perform the computations suggested by the theory. More particularly, we have reason for believing that they perform the computations in a strongly equivalent way; the algorithm employed is the same in both cases. This means we also have reason for thinking those creatures really do share the same psychological state types (as determined by the nature of the computational process). To use our theory of spoken word recognition for a quick example, we know that a human whose speech perception abilities are explained by the cohort model will token a phonetic representation in response to a particular kind of proximal stimulus. That particular representation—say, with the content '/r/'—is a psychological state type shared by any two individuals describable by the cohort model of spoken word recognition in response to the phoneme /r/. So, with the help of computational psychological theories, as opposed to trying to type mental states, we can more easily and confidently assert the psychological sameness of two individuals.

In the evidence I present later on to show that psychological state types are multiply realized, I will try to show how the flow of information in two particular creatures is the same from a psychological perspective in just the way outlined above. Clearly, however, there is more work to do. It is nice to have a way of evaluating the similarity of two creatures in terms of the realized states, but we also plainly need to know what the realizer state types are, too. In our case, that means figuring out what the brain state types are. That is the direction we have to go now.

Chapter 3: What is a Brain State Type?

Of course the strategy for dealing with psychological kinds will be commissioned again here with the task of sorting out the brain state types. In this case, I want to look at the kinds picked out by those sciences that deal with the brain. Fortunately, it is not as problematic as the mental side of things because there are not any 'intuitive' ways of marking the brain state types. Here we have no choice but to look at what the neurosciences are doing to settle the question of what makes for the same, or different, brain state types. So while there may have been a slight controversy about replacing 'mental' with 'psychological', there is not one in the case of talking about 'neuroscientific kinds' (or 'neural' or 'neurological', each of which I use interchangeably) instead of 'brain state kinds'.

But though this may be handy in one way, it introduces difficulties of another kind. The neurosciences, perhaps due to the fact that our knowledge of the brain has developed so fast, are not nearly as uniform in their treatment of what brain state types there are (compared with psychology). In fact, it is probably not right to say that neuroscientists are really concerned with typing brain *states*. Cognitive neuroscientists in particular seem rather to be concerned with correlating particular psychological functions with distinct brain *regions*—hence the prevalence of brain scanning and event-related potential studies. This constitutes a double complication. For one, it is not obvious how any region of the brain could count as a type of brain *state*. With some reason, one could complain that the former do not have anything to do with the latter. And if that is true, then it is not clear how an appeal to the sciences could help settle the issue between identity theorists and advocates of multiple realization. I will consider this problem later in Section 3.4. For another, there are so many ways that neuroscientists have proposed for carving up the brain into distinct regions. It is just not possible to inspect the brain casually and arrive at anything like a detailed map of the various regions. So it is no surprise to find there is a long history of attempts to decompose the brain to its canonical parts, some of which stuck and some of which were discarded immediately. For some, even today, the whole project of classifying the regions of the brain is a hopeless task. The main function of this chapter is thus to arrive at a demarcation of this mysterious 'map' of the brain.[5] As I said, there has been quite a lengthy discussion on just this subject, so it will be useful to trace this history in order to make sense of my view.

Before diving into the history directly, it might help to extend this analogy about map-making in order to explain the difficulties involved. With any map of natural terrain, there are bound to be many ways of setting up the borders. One way, of course, is to go with other natural features present in the environment. Islands, for instance, have quite obvious natural boundaries formed by the water enveloping the terrain. Similarly, rivers and lakes can do the same job for areas of land not surrounded by water. The Mississippi River and the Great Lakes are good illustrations. But natural boundaries do not constitute the only method of mapping. Many of the borders of the western United States, for example, are established by lines of latitude and longitude. Wyoming, Colorado, and Utah are states whose entire borders are established by longitudinal and latitudinal lines. So clearly cutting up the geography is not merely an 'intuitive' thing whose boundaries come pre-delineated. Sometimes drawing up the borders depends on abstract factors, or is at least influenced by abstract ideas. The brain is unfortunately quite

[5] Likening the process of individuating brain state types to making a 'map of the brain' is an analogy I shamelessly steal from Jennifer Mundale (see, e.g., her 1998). She is, so far as I know, the only philosopher to care much about chopping up the brain into distinct areas (and connecting the chopping with philosophical issues), so much of the following is influenced by, and indebted to, her works on the subject.

similar. There are some 'natural'-ish boundaries, like the interhemispheric fissure, the various prominent sulci and gyri (e.g., the lateral sulcus), and even the cerebellum, which is isolated much like an island at the back of the brain beneath the occipital cortex. But beyond some of these prominent 'landmarks' in the gross anatomy of the brain, the going is very much tougher in terms of breaking down the distinct areas. The cortex, in particular, is largely uniform and interconnected upon first sight. Though neuroscientists from as far back as the 19th century respected the difference between the four major lobes (frontal, temporal, parietal, and occipital), the appearance of each is indistinguishable; each lobe, though obviously separated spatially, is composed roughly of the same stuff scrunched up together (into the gyri and sulci) to save space. The situation is not much better with the brainstem either. There have come to be divisions in the brain stem between, say, the mesencephalon and metencephalon, but these were not the products of just rudimentary observation. The brain stem is actually physically continuous with the spinal cord. So, to say the least, figuring out which are the distinct brain states or areas is a very difficult question.

Despite the difficulties involved, however, I think it is possible to figure out a good map of the brain regions. Once the map is settled, then we can worry about how to connect those regions with types of brain states. Like I said earlier, in order to get things right about regions I think it is best to trace the history of brain cartography. This may seem odd given that reductionists typically frame their debate about completed, true sciences. All bets are off, for instance, about the reduction of current psychology to current neuroscience if the former and latter are woefully different from their finished, true counterparts. So one might think that looking into the history of brain science (when the knowledge available was vastly less than it is today) would be uninformative or irrelevant for determining the extent of the multiple realization of psychological kinds in brain state kinds. I want to drag through the history, anyway, for a few reasons. First, some of the historical ways of carving up the brain (including, of course, the one I want to adopt), are still acknowledged today. That is, they seem to have staying power that suggests that they will form at least some part of the completed neurosciences. Also, being clear on the various options will also help set the stage for the arguments I want to employ in favor of the taxonomy I adopt. Without knowing something about the various positions, my arguments might otherwise look misplaced and unpersuasive. Another important reason is that it will also allow me to straighten out some important confusions about the way in which brain mapping does or does not constrain the multiple realization of psychological kinds.

Section 3.1: Old and New Style Cartography

The most important historical distinction we have to keep in mind, drawn from the history of neuroscience, concerns the extent to which we can *localize* particular functions or psychological capacities to specific brain areas. There are some, on the one hand, who think that particular functions are localizable to some discrete area or other in the brain (perhaps not down to the single neuron, or a group of a few neurons, but at least to some chunk of the cortex—the degree, so to say, to which a function is localized is very much a relevant issue currently). On the other hand, there are some who think that psychological functions are not to be reduced to any single specific area in the brain—that, e.g., functions are typically distributed throughout the whole brain, so that all areas of the brain are equally important for, and involved in, psychological functions.[6] For convenience in what comes below, we can call proponents of the former view ‘localists’ and proponents of the latter view ‘holists’.

A few comments are worth pausing to mention here. First of all, this particular dispute, unlike the dispute about whether the brain is composed of discrete neurons or is seamlessly interconnected, is still actually quite active even today. The dominant view is the localist position, owing mainly to our increasing knowledge of the input systems (like vision, audition, or olfaction), whose processing is very much localized to particular areas of the brain—hence the names of certain areas of the cortex like primary *visual* cortex, supplementary *motor* area, etc. The modern holist has some lines of response, though. To begin, the brain is very highly connected—with intra-area connections (particularly feedback loops within regions) and inter-area connections that make, say, auditory information available to visual processing areas. To the extent that the brain is so interconnected, it appears to be a bit capricious to suggest that some discrete area is, say, *the* visual region. To extend this point, it also well known that there are ‘association’ areas in the brain (e.g., the auditory association area, or the visual association area), so named, clearly, because of the input from many different modalities. These areas are notoriously difficult for localizing the exact regions responsible for a given function.

Second, perhaps most importantly, despite the success of brain mapping in the case of the sense modalities, researchers have had virtually zero success locating the areas which support the more ‘central’ process (to use the language of Fodor, 1983). That is, there has never been an area in the brain discovered to be associated (at least exclusively) with abilities like problem

[6] Examples of localists are ubiquitous in the literature, since this is de facto the received view. For examples of holistic perspectives on brain function, see Uttal (2001) and Flourens (1824). Also, there is some reason to mark a distinction between holists and the so-called ‘equipotentialists’ (like Lashley, 1933), though for my purposes we don’t need to be so precise. See Mundale (2002) for discussion.

solving or decision-making. The frontal lobe, among other areas, is certainly heavily involved in processing of these kinds, but very little is known about how, by what mechanisms, and exactly where in the frontal lobe what work is being done. Our utter bewilderment at localizing these kinds of psychological capacities has been such a problem that even committed computationalists like Fodor (see his 1974, 2000, and 2008) have warned that such central processes could very well constitute exceptions to computationalist theories of psychological processes—or, more ominously, bring them down altogether.

So, with the difference between localism and holism under the belt, we can begin to look at a few of the classical attempts to individuate the distinct areas of the brain. We are not completely clear on all the problems that could be raised in this domain, but we have at least got enough of a foundation to survey some of the literature, from which, as it becomes necessary, we can raise these other concerns.

As far as I know, the first major attempt at locating distinct areas of the brain was undertaken by Franz Josef Gall and the other craniologists, or 'phrenologists' as they came to be called, around the turn of the 19th century.[7] This particular doctrine of phrenology had a few main, though very controversial, tenets. First of all, Gall believed that the brain was the 'organ' of the mind, an early statement of physicalism which at least the Catholic Church viewed with abhorrence. Secondly, he also thought the brain was broken down into sub-organs which specialized in particular capacities—e.g., a faculty for language, friendly attachment or fidelity,

---

[7] There were other historical precedents of the localization doctrine, most notably by Emanuel Swedenborg in his 'Oeconomia regni animalis', but also by writers like Thomas Willis of the 17th century who thought that memory and the will were controlled by the cerebral gyri, and even as far back as the Roman Galen, whose work set up the belief in the Middle Ages that psychological abilities were connected with activity in the ventricles (see Finger, 1994). But none of these early attempts at localization were nearly as systematic and extensive as Gall's treatment.

metaphysical perspicuity, and, my personal favorite, an impulse to propagation. The craniology bit, of course, is the result of Gall's measuring the cranium for bumps in order to see which particular organs were particularly well represented. This involved two more main assumptions: that the skull's shape accurately mirrored the actual shape of the cortex, and that the degree to which someone could claim, say, comparatively better metaphysical perspicuity was indicated in the bumps of the skull as an enlargement of the particular organ dedicated to that 'skill'. From these 'measurements', one could predict, for instance, what sorts of jobs an individual would be suitable for, what kind of personality someone would develop, or, worse, determine someone's mental aptitude.

Gall and the phrenologists were, of course, widely discredited—even, and for the most part, in their own day. And the discredit did not even derive from the commitment to physicalism. Many of the methods they used to confirm their hypotheses about the organs were not rigorous, and hence the evidence they cited was never taken very seriously by the scientific community. The effect, somewhat sadly, was temporarily to take the localist analyses down with them. Any position that advocated the localization of a psychological function was guilty by association—and, as one does these days with theories on which it is possible to pin verificationism, was argued away by reductio. It was not until the pioneering work done by Paul Broca in the 1860s on language processing that localist's view came back into fashion. The result was a debate which, as I said above, still actually has contemporary currency.

In the midst of this back and forth, Korbinian Brodmann developed his own mapping of the brain, which is still widely appealed to in current neuroscience (see his canonical 1909). The main thrust of his particular cartography of the brain was the claim that different structures inevitably correlate with different function. In particular, Brodmann is famous for having exhaustively studied the fine anatomy—viz., the *histology*—of the cerebral cortex. He noticed there were many clear structural and cellular differences across different areas. For instance, Brodmann noticed that the entire cortex is formed out of six layers of tissue. But the laminar profile of these layers compared across various locations of the cortex is vastly different. Certain layers, to provide an example, are input layers (information for that particular area to process come in at this particular layer) and therefore are thicker than the other layers in the area, and thicker than that same layer of adjacent areas. Also, a bit more basically, the cortex is not composed of the same kind of cells (neurons). There are, to name a few, pyramidal cells, basket cells, and Betz cells found throughout the cortex. One of the criteria by which Brodmann chopped up the brain was to note which areas were composed with which cells, or even which areas were composed of which combinations of cell types and the proportions in that area of each type. Among other things, Brodmann also looked at relative cell density, variation in cell shapes, and even affinities for stainings (Finger, 1994). Again, the point was to assume that histological variation tracked functional differentiation. So Brodmann was, in effect, advocating a thorough-going localist position. But his cytoarchitectural defense of localism also provides an excellent way of dividing up the different areas of the brain. Histological differences, that is, also track a plausible way of classifying the different brain state types.

Though the 52 areas into which he divided the cortex are still in use today—i.e., researchers still use these 'Brodmann numbers' to refer to particular areas of the cortex—Brodmann's findings were not without dispute in his day or thereafter. The main problem is that no one ever came to an agreement on which particular factors should be privileged in drawing up the boundaries. Should we, for instance, care more about the actual cellular composition of an area, or should we be more aware of the laminar profile of that piece of cortex? Early anatomists of the brain, predictably enough, differed in the precise boundaries they drew because of a lack of accepted principles for *applying* the kinds of criteria for delineating the brain areas. Karl Lashley, a famous holist and critic of Brodmann-style mapping and localization, published an important paper which attacked Brodmann's findings in just this way. In Lashley and Clark (1946), both men took brains from an Ateles monkey and provided maps of each according to cytoarchitectural features. Unsurprisingly, each map differed. Lashley also noted, however, that the differences were not entirely up to the caprice of the map-maker vis-à-vis the criteria employed. Some areas which were fairly clear in one brain were just simply non-existent in the other brain, suggesting that there is substantial variation in the (cellular/structural) organization of brains even among subjects of the same species. That this is the case, Lashley concluded, shows that it will be hard to localize psychological functions to discrete brain areas. But it also shows the inherent difficulties of demarcating brain areas by citing purely cytoarchitectural differences. Hard though it may be, it should be mentioned that Brodmann's maps were largely confirmed by subsequent findings about the localization of many functions, and it is still used today as a way of referring to different areas of the brain. Apparently his thought that cytoarchitectural differences tracked psychological specialization was at least roughly correct. Notice, finally, that his way of chopping up the brain does not depend on a localist construal. Since his map is cytoarchitectural, he could consistently have adopted holism if he wished. It would have undermined his guiding contention that function varies as a function of cellular composition, but he could have done so. Strictly speaking his map relies only on histology.

After Brodmann's anatomical mapping of the brain, we can look at more modern methods of typing different brain areas. These methods all share the same assumption that individuating distinct areas is easiest when we start with *functional localization*. That is, start with a localist notion of function, locate the function you are interested in via one of the following methods, then just call whatever bit you locate a distinct brain area. For example, probably the main criterion used today is to trace the connectivity of certain areas of the brain. If, for instance, a piece of the cortex is found to connect ultimately with the retina, via a few intermediate stops in other brain structures, then odds are that area is concerned with processing visual information—and, according to our guiding assumption, therefore composes its own distinct region. In some instances it is also profitable to look at topographic layouts in certain areas. The organization of the neurons in the primary visual cortex, for instance, are laid out in a retinotopic manner, with certain groups of cells exclusively devoted to processing information from individual receptors on the retina. That is, different cells are concerned with handling the information present in a (sometimes) very tiny area of your visual field. This organization of receptors in the retina, in fact, is faithfully mimicked in V1. The story is the same for somatic information in S1; the different areas of the body (hands, feet, thighs, etc.) each have a different bit of cortex devoted to keeping track of the information coming from each area. So, too, the primary motor cortex: it has a topographic map of the body where the different bits of cortex control different limbs (right down, for instance, to the level of fingers). So one way of localizing functions to some discrete area of the brain is by checking for neural arrangements like these which make tracking information about particular parts of the body easy—again, the point being that noting functional dedication is a handy way of mapping the distinct areas of the brain. Finally, PET scans and fMRIs also contribute to this process of functional localization. In particular, these methods track metabolic activity in the brain. This higher metabolic activity, in turn, is assumed to track the areas of the brain which are performing the functional tasks. So if you have someone perform a particular visual task, you need only look for the higher levels of metabolic activity to establish which areas process visual information. These methods are somewhat crude still; the spatial resolution for fMRIs, though quite good, still is not exact enough to arrive at clearly defined boundaries for certain functionally important regions. Or, at any rate, those who are interested in mapping the brain probably would not find these scans terribly helpful. But when combined with other methods (e.g., noting topographic organization), it becomes possible to arrive at significantly more precise boundaries.

Section 3.2: Taking Stock of the Options

So that is a brief enough look at some of the history for sorting out the different areas of the brain. I want to pause now and make a few comments about the foregoing which will bear on the kind of cartography I will endorse below, and perhaps will also clarify further some of the concepts I have been invoking. First of all, I hope that it is clear how contentious this issue about typing brain areas is. The multitude of criteria we can bring to bear inevitably leads to a blurriness of the boundaries of all the distinct areas. Any position I wish to adopt, then, had better come on the back of at least some argumentation. I have got to say something about why we should rule out *this* method of cartography rather than *that* method, why, say, using purely cytoarchitectural criteria is preferable to using purely functional criteria. This is the major goal in what remains of the chapter.

Secondly, extending this first point somewhat, notice that advocates of multiple realization and identity theory are not obliged to make any stands on the positions we have discussed above. Both positions make an essential reference to brain state types, so in either case we will have to say something about what those brain state types are. It might seem natural for proponents of both views to follow the localists in using functional considerations to guide the creation of all the borders of the different areas. But such a move is not obligatory given their other commitments. And this is actually a very good thing (see the section below for why). Localism seems particularly important for brain mapping only because the view also provides a quick, obvious means for adjudicating boundaries. Holism, on the other hand, only wishes to maintain something like the claim that psychological functions are not carried out by any singular part of the brain—and so by extension, I suppose, by any bona fide brain region. But that does not matter. Quite consistently, the holist can allow there to be as fine-grained a taxonomy of brain regions as you like (say, the brain is composed of 2, 348 distinct areas), or an incredibly coarse-grained taxonomy (say, there is only one distinct brain area: the brain itself)[8]; he must only hold that none of those areas is privileged with respect to some psychological capacity. Perhaps the identity theorist would be forced, *qua* holist, to adopt a rather coarse-grained taxonomy, because the more brain areas you posit the more the distribution of the processes will inevitably lead to particular psychological kinds being realized in a multitude of brain state kinds. But an identity theorist could, in this way, retain his commitment to holism. Clearly the localist position could be adopted by the identity theorist or nonreductive physicalist, too. In this case, the boundaries for the areas are drawn according to functional considerations. But that is entirely consistent with the further fight between the views about whether the functions localized to a particular area are, as it were, essentially tied to those areas—or whether, e.g., those same functions can be carried out in different areas across different species, within the same species, or perhaps even within the same individual. Nothing about this latter dispute

[8] The localist, we can mention in passing, is conceivably also in this boat: she could also have a taxonomy so rich or so sparse that it outstrips or underdetermines the amount of functions she also posits. I have no idea why a localist would do so, however—and certainly many classical localists typed brain states strictly by correlation with psychological functions.

impugns the localist's claim that psychological capacities are realized in particular local areas, however.

Thirdly, to build even more on the foregoing, we must remember that, though brain mapping traditionally has been importantly linked with the project of localizing functions, the prospects for the former do not rise and fall with those of the latter. Jennifer Mundale might be right in saying that "[i]f brain operation is not determined by local differential contributions of parts, but by more holistic principles, the enterprise of brain mapping will be much less significant for cognitive science" (Mundale, 1998, pg. 139). But significance for cognitive science, whatever that comes to, hardly mandates a functionally-minded cartography of the brain. There could be other considerations to bear in mind, beyond the issues inherent in the localism/holism debate, while drawing all the lines. So we should be very careful not to assimilate the constraints on localist views about functions with the constraints on those who are concerned with making a map of the brain. Drawing the boundaries through functional lenses is a particular method one could adopt, but it is certainly not the only one. The various sulci and fissures which are still used to indicate different areas constitute obvious counterexamples.

Section 3.3: Why Brodmann's Areas are the Best Bet

So what are we to do? We have got a lot of options. We could stick doggedly with the 'natural' landmarks, treating as distinct brain areas those locations that are separated by fissures and sulci, or by being disconnected from each other (like the cerebellum). We can go with Brodmann and privilege cytoarchitectural features of brains as the real border-defining factors. Or we could go in for very functionally-oriented localism and, in particular, allow our insistence that psychological capacities are isolated to and realized in distinct locations in the brain to pull double duty and provide us with the borders of the various brain areas. Or we could try to find some sort of hybrid position between the above views, or maybe develop one of them by incorporating some of the more modern techniques (like taking into account connections between brain tissues).

A final bit of commentary will perhaps elucidate the difficult nature of answering this question. What would be ideal for people worried about identity theory and multiple realization is if the mapping of the brain were done from the perspective of a purely brain-oriented science. That is, we would like it to be the case that there are interesting generalities to state about the brain without having to worry about other kinds posited by other theories—in particular, *psychological* theories. The identity theorist and the advocate of multiple realization are both trying to examine the relationship between two different levels of analysis: the brain science level and the psychological level. The assumption is that, after we articulate each level independently, we will notice that there is, just as a matter of fact, either a one-to-one or one-to-many mapping of the psychological kinds to the brain science kinds. If one *only* incorporates the psychological level in figuring out the kinds (in this case, the areas) at the brain science level, then the well is simply poisoned in favor of identity theory from the beginning. Then trivially there will be a one-to-one mapping of psychological kinds to brain science kinds because the latter are defined with reference to the former; the 'pain area', to make up a brain state kind, will just be any particular bit of the brain that makes us feel the pain we do. Even if we were to ablate that 'pain area' and, by some kind of plastic effect, one retained the feeling of pain in some (spatially) different bit of the brain, that new bit would still constitute the same 'pain area' because of our reliance on psychological functions to shape the boundaries of the brain regions.

That new part of the brain which carried out the 'pain-feeling' processes would simply become the new, but taxonomically equivalent, 'pain area'. So even in this case, where the psychological level appears to cross cut the brain level in at least roughly the way multiple realization would anticipate (where the psychological kinds can be realized in just about whatever bit of the brain you like), identity theory would still carry the day.

The tradition of localism is probably the received view these days, and researchers sometimes do use functional criteria to help separate distinct regions in the brain.[9] But history has also had influence in current brain mapping through Brodmann's cytoarchitectural investigations in the early twentieth century. Brodmann was, of course, a localist in that he thought that different psychological abilities would associate with histologically different areas of the cortex. But functional considerations—like connectivity analysis, topographic organization, or PET scan studies—had no effect on his kind of cartography. The instruments needed to employ those sorts of methods did not even exist until decades after Brodmann. Instead, he divided up the cortex by examining the differences in the cellular composition of the cortex, and then later researchers were able largely to confirm the functional significance of the brain regions with a unique cytoarchitecture. Brodmann's map, then, avoids this problem about typing brain regions by only functional criteria. He cared about localization of function, but his technique of brain mapping never appealed to any functions to do the work.

Thus I am suggesting that we should rely on Brodmann's mapping of the cortex to constitute the taxonomy we will appeal to in establishing sameness of brain state type. The main reason is the issue I have raised about trivializing identity theory. In order to evaluate this view (in contrast to multiple realization), we need the psychological level and the brain level to be

[9] See Bechtel and Mundale (1999) for much more in this vein. In Chapter 5, sections 4 and 5, this will become an important issue to discuss.

distinct. The basic idea behind identity theory is that the same kind at one level of abstraction (say, the psychological level) is realized in one particular (homogenous) kind at another lower level of abstraction (say, the brain science level). The work the respective sciences have to do in order to articulate their proprietary kinds, however, is supposed to be independent. That independence is precisely what gives bite to the claim that psychology is not autonomous from, for example, neuroscience: that, when the sciences cut up the world according to the types that figure into their laws and generalizations, we happen to find that the higher level kinds just give different names to events which are *already* invoked by the lower level theory. The identity theory, if true, is thus supposed to be an interesting and substantive empirical outcome; *it is not supposed to be the petty consequence of individuating kinds by reference to another science's kinds in a one-to-one fashion*. So if brain science identifies its kinds by reference to psychological kinds and functions, then it appears that it is just a trivial, conceptual point that psychological kinds are realized in singular brain state kinds and, further, that psychology is after all totally autonomous from any considerations about the brain. So the identity theorist would, in a sense, win the battle by insisting on only functional criteria in mapping brain states, but the war would be well lost.

Another consideration to mention here is that the original advocates of identity theory did propose their theory as a way of relating whatever state types a science of the mental would adopt with whatever state types a science of the brain would adopt. The analogy with the reduction of heat to mean molecular kinetic energy makes this point. The assumption was that this particular way of conceptualizing the world (by doing science A) picks out the exact same things as another, distinct way of conceptualizing the world (by doing science B), with preference given to science which was privileged in whatever important ways—e.g., by being more unified with other sciences, or by explaining more phenomena than the compared science. So, the early identity theorists were basically offering an educated guess about what the future would say about the nature of the mind. But, whatever the outcome, it was always meant to be a discovered, empirical fact. To that extent, it is not even clear, if we were to require brain state types to be individuated by location of psychological processes and kinds, that the resulting view would really *be* the identity theory.

The second reason in favor of adopting Brodmann areas as identifying the brain state kinds is that these areas still have huge currency in contemporary neuroscience. As with typing psychological kinds, the best bet is to let the scientific theories work themselves out. Whatever remains will automatically provide the ontology that is best justified. That neuroscientists still use Brodmann areas to refer to specific regions of the cortex is great reason for thinking that those areas describe genuine kinds of brain states. Recall earlier that we had worried about relying on outdated methods for chopping up the distinct parts of the brain—because evaluating the prospects of multiple realization ultimately depends on relationship of *true* psychology and brain science. There is prima facie reason to complain about Brodmann areas on these grounds, but noticing the reliance of current neuroscientific research on invoking these areas should allay any such fears. Although his work is definitely old, Brodmann areas are still neuroscientifically legitimate. Combine this point with the perils of adopting mapping criteria that are too functionally-minded, and the case seems rather decisive in favor of going with Brodmann's map.

There is one kind of drawback for the adoption of Brodmann's map—viz., that it is only a map of the *cerebral cortex*. The human brain is, however, composed of much more tissue. The cortex is the tissue we probably most easily associate with the brain because this part is folded around everything else (but the cerebellum). All of the thalamic areas, the midbrain, down to the spinal cord, though covered by the cortex, remain completely unmapped by Brodmann's scheme. These subcortical areas account for a pretty large amount of the tissue of the brain—and, of course, a good amount of the functions which the brain carries out. Most of the essential functions for survival, like autonomic responses, control of heart rate, breathing rate, and sleep are taken care of by these structures. And the thalamus is incredibly important for organizing and distributing a great deal of sensory information to the appropriate cortical areas. So one might legitimately complain that Brodmann's studies are inadequate for brain state individuation because they are incomplete.

I take it that this worry can be overcome, however, simply by maintaining that the true criteria by which different brain state types are identified has mainly to do with cytoarchitectural features. Just because Brodmann never mapped the subcortical structure does not mean the principle by which he distinguished the different brain areas is somehow compromised. Probably the reason why no one has bothered to type these structures along Brodmann lines is because they lie so deep within the brain (and that they perform the more primitive, less interesting functions), not because it would be impossible to provide such a mapping.

Section 3.4: Why do Neural Areas Provide a Good Guide for Brain State Types?

Recall the first complication I talked about back in the introduction to the chapter. There might be some reason for objection here that, strictly speaking, Brodmann's areas are just brain *regions*, not necessarily brain *states*. And though I have tried to give reasons for why typing brain states along Broca's lines is the best bet, the identity theorist might object that the emphasis I place on those areas simply misses the point. More particularly, the identity theorist could complain that her identities have nothing necessarily to do with brain regions or precise areas of neural tissue, but rather just involve an identity between types of psychological and neurological states. Some lump of the brain, or some given pattern of cellular features, according to this move, might have no essential connection to the real taxonomy of neurological state types. The identity theorist might therefore conclude that the kind of multiple realization I am after only undermines a conclusion she never held in the first place.

There are, I think, two ways of understanding this objection that merit discussion. First of all, one could read the objection as being that the *whole* story about brain state types cannot be told simply by appeal to Brodmann's areas (or any other notion of brain region). On their own, Brodmann's areas are just that: regions of the cortex. We would have to add in something about *a pattern of neural activation in some Brodmann area* to tell the full story about what a brain state is. I think something like a pattern of neural activity *and* that activity taking place in some Brodmann area are both jointly sufficient for something to count as a type of brain state, but I will not really try to argue for that claim. No one really knows the precise details about neural firings and what psychological consequences they carry, so it is difficult to specify the exact sufficient conditions for some individual belonging to a type of brain state. But this is not an issue for me to be too concerned about. What really matters is the necessary condition about some individual counting as a brain state type only if it is associated with a particular Brodmann area. The identity theorist is the one who has the burden for coming up with the complete analysis. She has to explain exactly what it is that is identical to this psychological kind. The evidence I present will at least violate this necessary condition, showing that various psychological kinds are realized in neural regions that are typed as distinct according to

Brodmann's areas, and, therefore, as different brain state types. The rest of the story is one I do not have to tell.

The second way of reading the identity theorist's objection is as a sort of blanket response to *any* attempt at brain state type individuation along the lines of brain regions. I argued, in favor of Brodmann's areas, that they at least have neuroscientific validity even today—therefore satisfying my desire to have the issue of multiple realization settled by the taxonomies of the sciences in question. It is also important that the areas are orthogonal to functional/psychological factors, principally for using the areas to evaluate the multiple realization of psychological kinds. But the identity theorist might find herself still unimpressed, questioning why it is that regions of the brain should even be necessary conditions on the individuation of brain state types.

In response to this line, I think it is important to point out that all of the empirical work that identity theorists could use as support for the view is framed in terms of activity in some neural region or other. That is, a very plausible way of spelling out the relevant identities would look very much like '(visual) edge detection is identical with such-and-such activation occurring in such-and-such a Brodmann area'. The 'such-and-such activation in such-and-such a place' leaves an obvious promissory note, but not one to raise any important concerns. As the neurosciences progress, it will become possible to state the identity theorist's thesis more accurately.

In the meantime, it is clear that a plausible source of the identity theorist's identities will invoke brain state types that are defined by neural areas like those picked out by Broca's famous areas. This notion of how to construe the identities—and hence the brain state types—along the lines of neurological areas is valuable for a few reasons. First of all, as I mentioned above, it provides a scientifically-oriented notion of brain state kinds. This is important for matching the historical background from which the identity theory and multiple realization were born. But it also does so in a way that does not trivialize the identity theory. It is still left open for discovery that the same psychological functions are instantiated in activations in all the same Broca's areas. Second, it provides definition to the identity theorist's view, making it clear exactly what is supposed to be identical to what. This is important for the obvious fact that the identity theorist has no view if she does not make it clear exactly what the brain state types are—making the view impossible to attack and also to defend. Finally, it also introduces some empirical respectability for the view. Of course, a major goal of the dissertation is to undermine this support, but initially the results of today's popular brain scans (like fMRI and PET) in correlating neural activity with specific brain regions leave the identity theory in good standing. I take it these are also strong reasons for me, the identity theorist, and everyone else to take seriously the proposal that brain state types are best understood as neural activations within Brodmann's areas.

So the important part of this chapter is that I argue that a sufficient condition for two different brain states x and y to be of different types is that x and y must have different cytoarchitectonics (they must be classified by distinct Brodmann areas). The main implication of this condition is that two individual brain states may not automatically be grouped together in the same class simply because they perform the same cognitive task. This is important in the context of multiple realization because it opens up the possibility that the same psychological function is carried out in genuinely different types of brain states.

Now that we have got both taxonomies in hand (of psychological states and brain states)—or at least we have got outlines of how to figure out the appropriate psychological taxonomy—we stand in a position from which to evaluate the prospects of multiple realization.

All we have to do now is identify the psychological kinds and see how they match up with the Brodmann areas. If it turns out that the same psychological state type is correlated with many different Brodmann areas, we will have strong reason to infer multiple realization. That is, we will have good reason for thinking that the same psychological kind is realized in different brain state kinds. From there, we also have strong reason to reject the identity theory.

Chapter 4: What Count as *Multiple* Realizations? A Reply to Shapiro

Before moving on to the empirical evidence for multiple realization, I would prefer to remove some of the conceptual criticisms leveled at the argument from multiple realization. To begin, I want to look at Shapiro's attack on the notion of *multiple* realizations. For whatever reasons, there has not been very much literature on this particular topic either. Aside from the realization relation, the story about how to classify the relevant kinds, the problem about how to tell apart multiple realizations is the only other major part of a multiple realization claim. Despite the lack of interest, I do think the work that Larwrence Shapiro has done on what makes realizations different *does* substantially affect my view. And I do think some sort of response is necessary to the criticism he would have of the evidence I want to offer in favor of multiple realization. Here, in other words, is a place where there is substantive disagreement between my view and his from the very start. If I cannot provide some remarks about why his view is mistaken (and that my view is, in some way, superior), then my project never gets off the ground. As we go along, I will explain this point about why a defense is necessary on my part.

Section 4.1: A Short Explanation of Shapiro's Attack

Let us first rehearse what he has to say about multiple realizations (to repeat an earlier citation, see especially his 2000 and 2004), then we will move on to the evaluative part. Shapiro's main point is that the empirical evidence one requires in order to establish multiple realization (he calls it the 'multiple realizability thesis') is much harder to come by than typically thought. The average evidence given in support of the thesis involves appeals to intuitions about the possibility of creating a mind out of Swiss cheese, or the possibility of artificial intelligence, or even extraterrestrial intelligence. Beyond this simple appeal to intuitions, however, the rest of the evidence about the empirical likelihood of multiple realization is either assumed, or taken to be unnecessary. Shapiro does not find this kind of evidence very convincing; resting the case for multiple realization on such intuitions and empirical assumptions is not going to cut it. So his goal (at least in his 2000) is not to argue for a reductive view of mind as much as it is to show that the case for multiple realization is not a lock.

He does so by scrutinizing what it takes for realizations of some kind to be *different*. He takes it as his starting point that for all realizers there are properties which are relevant to the realization and properties that are irrelevant to it. The trick is to figure out which are the properties that are relevant. Shapiro wants to show that it is not as easy as intuition makes it out to be. One general concern is that we will group things as similar and different based on the level of description we are using to individuate kinds (cf., Bechtel and Mundale, 1999, pg. 203). If we are being suitably vague (if our level of description is a bit 'higher'), then we might think it is fine to group anything that can *see* into the same psychological category—completely neglecting the fact that many different creatures employ many different methods for their visual abilities. At this level of description, we would only be concerned with, roughly, using information about the environment carried by electromagnetic waves to keep the organism alive. The differences in the eyes, number of eyes, lenses, visual processing, and all the rest would not matter, because we would only be concerned with grouping together all things with this psychological capacity. From a more focused level of description (from a 'lower' level), however, we would not group all sorts of vision together. We might, for instance, care about the precise details of a creature's visual abilities. Or, if we are really focused, we might worry about, say, the differences in the number of constituent atoms in different retinas. There is bound to be a difference in all retinas if we describe things at that level. What kinds of samenesses and differences are important, however, for the appraisal of multiple realization? Is it not problematic that one could easily adopt a very coarse-grained level of description for individuating realized properties, while adopting a very fine-grained level of description for individuating realizer properties? Surely supporting multiple realization in that way is cheating.

To use another example that illustrates the worry, which Shapiro likes to use, think about the case of two corkscrews that differ only in color. We have two realizations of the 'kind' corkscrew, but do we have two *distinct* realizations? Could we say that these two token corkscrews count as evidence that the type/kind 'corkscrew' is multiply realized? Shapiro—and everyone else, it seems—claims that we do not, that two corkscrews differing only in their color do not count as different realizations. So, at least in this case, the colors of the respective corkscrews count as properties which are irrelevant to the realization of the corkscrew. That is fine, but it obviously begs some important questions. How can we tell whether two realizations of some kind are distinct realizations? What level of description should we adopt when making these kinds of judgments (or, in other words: what sorts of differences and samenesses are the ones we should consider)?

A good starting place for answering these questions, Shapiro thinks, is to begin with kinds that we think have a good chance of being multiply realized. In particular, he thinks we should look at *functional* kinds. We should look at those types of things that are typed by what they do. For example, the typical nonreductive physicalist loves to talk about mousetraps, carburetors, and computers, because it is at least possible that these sorts of things, defined as they are by the operations and capacities they perform, can have many different kinds of occupants fill the roles specified by those operations and capacities. We sort objects into the kind ‘mousetrap’, for example, by those objects’ ability to trap mice. Whatever object can get the function ‘live mouse in, dead mouse out’ right will be sorted into the ‘mousetrap’ kind. Given this sort of idea, we have a natural suggestion for what properties are the ones that are relevant for realizing the functional kind in question: those properties are the ones that are responsible for bringing about the function. Of course the same similarity problems arise for identifying what function of a corkscrew is interesting to us. Probably we are concerned with some corkscrew’s ability to remove corks, but we could also be interested in the corkscrew’s ability to make a jingling sound. But what is important is that, once we have fixed what sort of function it is that we will concern ourselves with, there is a fairly objective functional analysis to be made that can identify the properties responsible for the jingling (or whatever). Such objectivity is welcome in response to the usual worries about judgments of similarity and distinctness. In particular, functional kinds make it very easy to pick out the function we want to analyze: whatever function it is that individuates the kind. So there is all the more reason to focus on functional kinds. With respect to the corkscrews, for instance, we can say that the properties of having such-and-such a color is not part of the realization of the corkscrew because those properties have nothing to do with what makes the object a corkscrew (what makes it perform the function of removing corks). Whether it is blue, red, or green, that sort of consideration does not matter for whether an object can get the corkscrew job done or not.

So now we have a nice way of sorting out the relevant from irrelevant properties when figuring out the realization base for some functional kind. And this is good because Shapiro thinks this ability to sort the relevant from irrelevant properties via functional analysis provides us with a straightforward answer to questions about which realizations are really distinct, too.

That is, in the same way that worrying about which properties are the realizers for functional kinds leaves us wondering about which properties make the function possible, so too worrying about what different ways there are of performing the same function will provide us with a guide for figuring out which realizations are distinct. For Shapiro thinks that two realizations of some function count as distinct when those realizations bring about the function in *different ways*. Or, to put the point differently (Shapiro, 2000, pg 644): "if two particulars differ only in properties that do not in any way affect the achievement of the defining capacity of a kind, then there is no reason to say that they are tokens of different realizations of the kind".

This sort or remark immediately makes sense of our intuition that two corkscrews that differ only in color do not count as distinct realizations. Those properties—the color of the corkscrews—have nothing to do with the actual removal of corks. It is rather other properties like the rigidity and shape of the corkscrews that account for the ability to do the things that type them as corkscrews (viz., remove corks). On the other hand, this remark also makes sense of the intuition we have that the waiter's corkscrew and the winged corkscrew *are* in fact distinct realizations of that kind. Since the former removes corks by using a lever, and the latter removes corks by virtue of its rack and pinion, they clearly perform the function individuative of the kind in different ways. Shapiro's claim is that performing the same function in different ways is what counts as the criterion which distinguishes distinct realizations of a kind. That is, what makes multiple realizations genuinely *multiple* realizations is they differ in their causally relevant properties.

But here is the rub. Notice that what Shapiro is claiming here affects the case of two particular, say, waiter's corkscrews which differ only in their *material composition*. Imagine one of the corkscrews is made of steel, the other of aluminum. Of course our Fodorian intuitions seem to tell us that we have a paradigmatic case of multiple realization (of corkscrews). Differences in composition seem to show that the actual constitution of the function is always important for evaluating multiple realization claims—exactly the kind of point advocates of the autonomy of psychology have been at pains to show about psychological processes. So the function in question is multiply realized when it can be instantiated in materially different substrates. But, by Shapiro's lights, we would have to withdraw this claim in favor of the judgment that the two corkscrews were not genuinely *multiple* realizations. Why? Because (at least in this corkscrew case) the differences in the material composition to do not realize the function of removing corks in different *ways*. The corkscrews remove corks because, roughly speaking, they have the property of rigidity; it is because the corkscrews are strong and not malleable that they are useful for digging corks out of bottles. The steel and aluminum composition, of course, are relevant for realizing the rigidity in each corkscrew—but, critically, they both realize *the same functionally relevant property*. When we analyze each device, we see they have the same functional decomposition. In each object, the same properties are used to make them cork-removers. So, according to Shapiro's criterion about individuating distinct realizations, two token corkscrews made of different materials, assuming those corkscrews perform the function of removing corks in the same way, are no more *distinct* realizations than are two token corkscrews which differ only in color. Being made of aluminum or Swiss cheese or whatever, holding everything else equal, is not enough to establish that a kind is multiply realized.

This is why Shapiro begins by saying that his goal is not to show that, say, psychology is reducible to neuroscience, rather that finding the empirical evidence you would need in order to show that psychological processes are multiply realized is going to be much more difficult than previously thought. Off-hand references to the possibility of constructing a mind out of cheese, for instance, do not even come close. In fact, one could not (just) make references to the progress made in artificial intelligence. Notice that what would be necessary is to show that artificial intelligences realize their capacities in ways that *differ* from how our brains realize the same capacities in us. This is partially what Shapiro means when he makes the somewhat astonishing claim (Shapiro, 2000, pg. 645) that we could replace every neuron in your brain, one by one, with a silicon chip that performs just the same process (in just the same way) that the replaced neuron served and we would still *not* have a distinct realization of a mind. Imagine, then, the task that confronts the person who thinks that psychology is autonomous from the 'hard' sciences (i.e., is multiply realized): she must first figure out how it is that we perform the psychological functions we do, then show that those same functions are accomplished in other creatures (or machines) in substantially different ways. She would, that is, have to understand fully how it is that humans process visual information, then show how some other creature (or machine) realizes vision (that is much like ours) in a functionally different way (e.g., the information gets processed in a different manner). Obviously, this is not going to be an easy task—*particularly* as the level of analysis becomes more fine-grained. The case for multiple realization is therefore going to be much harder to establish than philosophers of mind have generally tended to think.

Fine, so what does all this have to do with my dissertation? Why should I confront this view about distinguishing truly *multiple* realizations? There are two reasons. The main reason is simply that it presents a burden on me that I do not believe I have to meet. Namely, I want to produce examples of genuine cases of multiply realized psychological kinds—but I do not think in order to do that I have to meet his criterion about finding differences in realizations that make a difference in how the all-important function is performed. So if I want to convince the readers that the examples I adduce are bona fide cases of multiple realization, it would be nice if I could say something about why Shapiro is wrong to make the demand he does. The second reason, perhaps more importantly, is that I do not think that I can consistently hold (a) that a suitable way of establishing sameness of psychological type is by appeal to strong equivalence of the process, and (b) that the proper way of figuring out which realizations are distinct is by appeal to differences in how the functions are carried out, and (c) that there really are legitimate cases of multiple realization. For in securing (a), it seems like I automatically guarantee that I cannot satisfy (b). Once you establish that two processes are algorithmically equivalent, that would seem to show that the realizations are not brought about in functionally different ways. The equivalence of the processes makes it secure that, from a functional standpoint, the realizations are the same—that, for example, the process we are concerned with is accomplished via the same effective method. So in holding (a) and (b), I would make it impossible to hold (c), which is precisely the opposite of what I want to say. Since I also happen to believe that (a) is the best, though admittedly quite stringent, way of establishing psychological sameness, I have to deny (b) in order to support the view I am ultimately after.

So it seems I am on the hook for showing why Shapiro is wrong. What sort of reasons could there be for denying his view? What sort of alternate views on the matter could one adopt? Let us begin with the second question first. I will offer an alternative account of what accounts for two realizations being truly distinct, argue for it, then show what sort of effects my view has on Shapiro's. Hopefully, after laying out what I think accounts for genuinely distinct realizations, the reader will think I also have strong reasons for denying Shapiro's own account.

Section 4.2: Arguing for a Different Account of Distinct Realizations

This alternative will not be surprising given what I have said already about brain state types. To put my proposal into a sentence: what matters for whether or not two realizations of a mind are distinct realizations is a function of what the scientific theories at the realization level (the brain science level) are saying. So, much like I maintain that the obvious way of establishing sameness of psychological state type is by simply looking at the prevailing psychological theories, so, too, we must appeal to the work of those who care about the brain in order to understand what constitute genuine differences at the realizer level. In particular, we have to look at the taxonomies of the scientific theories in order to determine whether we are dealing with genuinely distinct kinds. If, in order to articulate its laws and generalizations, a theory has to postulate different kinds of things to explain all of the relevant phenomena, we have a ready-made method for separating distinct realizations: they are just the realization bases which are *distinct kinds* (from the perspective of the theory which makes use of the kinds). This deflates the worry about distinct realizations to worrying about which kinds are different. But that worry is no problem. The different predicates of the science pick out its different kinds. So all we have to do in order to see whether psychological state types are differently—genuinely multiply—realized is to take a look at what state types they are realized in, looking for whether those state types are different (according to the taxonomy of the brain science). If we see the same psychological kind realized in kinds that the brain science individuates as distinct, then it must be the case that that psychological kind is really multiply realized.

I am not sure if this proposal will work generally speaking; I do not know if *all* questions of whether some type of thing is multiply realized can be resolved by appeals to scientific theories at the realizer levels (to say nothing of the question of such appeals at the realiz*ed* level for all types/properties). In fact, it seems pretty clear that they cannot be. Take, for instance, the case of the different varieties of corkscrew. I suppose there are many—Shapiro included—who want to hold that the waiter's and winged corkscrews count as distinct realizations of the type corkscrew. That is fine, but I do not imagine there is any science, or scientific theory, that ranges over these different realizations and classifies them differently, e.g., for its explanatory or predictive purposes. So unless I want to have an account of multiple realization which rules out such cases like the corkscrew, or cases involving mousetraps, I would have to admit my proposal applies only to scientific kinds.

I think that such a consequence is okay, however. It seems to me that in discussions of multiple realization we have really got two different notions typically at play. Sorting out these two general 'senses' of multiple realization will help explain the view I want to advocate and might help me more easily explain what I find objectionable about Shapiro's. On the one hand, then, there is a scientifically respectable sense of multiple realization which has to do with the comparison of kinds across scientific theories. The relative precision of this sense of multiple realization is the result of the precision of the sciences. The practices and standards from and by which the legitimacy of a taxonomy of some science is derived act as independent constraints—constraints which give strength to claims about some particular kind being multiply realized by others. We do not simply get to make generalizations about intuitions we have from particular cases in order to state general principles about what it is for some scientific kind to be multiply realized. Scientific kinds are independently established by the sciences which use them for their predictions, explanations, and generalizations, and philosophers merely come along later to check on the relationships which hold between the different taxonomies (viz., whether a kind of the higher-level science is correlated with distinct kinds specified by the lower-level science).

On the other hand, we have the loose, more ambiguous, sense of multiple realization that applies to non-scientific kinds. The history of philosophy of mind is littered with these examples (Shapiro included): corkscrews, mousetraps, carburetors, etc. Here the examples do not have any kind of independent checks on what constitutes a genuine kind, particularly at the realizer level. We may have a somewhat clear, though obviously unscientific, understanding of what kind of thing a mousetrap is (e.g., anything that instantiates the function 'live mouse in, dead mouse out'), but all bets are off when we start looking at the actual contraptions in which the mousetrap is realized. Here there is no clear method for deciding when the realizations are of the same or different kinds—whether, for instance, merely compositional details should make a difference, or whether we should focus on different means, as it were, of instantiating that mousetrap function. Compare the case of carburetors: we know they mix fuel and air, but we do not know if we should count as distinct those which have one or multiple venturis.[10]

I take it that exactly this ambiguity of the realizations bases of the non-scientific kinds is the root of Shapiro's concern. Since no one has ever given us any reason to prefer one 'taxonomic' standard to another when it comes to evaluating these non-scientific realization bases, Shapiro is going to make a suggestion (or argue for that suggestion—see more below on how he arrives at his principle for distinguishing distinct realizations) about how to do it. My reply at this point is to say that it does not really matter whether we can come to a more specific sense of when a non-scientific kind is multiply realized. As far as I can tell, these sorts of examples have been used mostly (if not entirely) as aids for understanding the more rigorous,

---

[10] Though I am focusing here on distinct realizations, the same worry of ambiguity infects the sameness of the realized kinds. It is not totally obvious that all the various devices that pass for mousetraps ought to be classified together. Why should someone not say that spring-loaded mousetraps are of a different type than sticky-floored ones? (See Funkhouser, 2007)

scientific examples. Trying to carve out some principle that captures our intuitions about these ambiguous cases is like trying to codify what it is that people typically mean when they talk about fruit. Even if we could come to some understanding about that (for the folk, fruits are supposed to be sweet non-candies, for instance), we would not have made any kind of important progress. There is still a botanic distinction between fruits and non-fruits which takes obvious precedent. In fact, popularizing what the non-expert thinks is likely only to cause confusion. I think something like this is going on with Shapiro's account of what makes two realizations distinct. He is suggesting a principle that he hopes will give us some way of settling these ambiguous cases, but it will obviously face problems when it comes to cases where what makes two kinds distinct is already established (viz., in cases involving scientific kinds). We already know what it is for scientific kinds to be multiply realized, and there is no need to worry when that precise sense of multiple realization fails to help sort out the ambiguous cases. It is far better to let those cases remain imprecise rather than allowing that imprecision to infect the well-defined cases—particularly when the ultimate question is about the multiple realization of psychological kinds (which are of course scientific kinds).[11]

That constitutes the main reason why I can insist on using the more strict sense of multiple realization: its precision offers a rigorous framework from which to see whether psychological kinds are multiply realized in neurological kinds. But there is another important reason as well: the history of multiple realization and the identity theory is based in the debates

[11] It is worth pointing out here that the same lack of precision infects Polger's different varieties of multiple realizability. Each of his four notions (e.g., Standard-MR, SETI-MR, etc.) simply relies on some intuitive notion of how multiple realization works. If what matters for truly multiple realizations depends on scientific taxonomy, there is no reason to distinguish between slightly, significantly, or vastly different realization bases. I take it this criticism also means that his strategy for dealing with multiple realization—his arguments for why some varieties are accommodated by identity theory and some are abandoned as question-begging—will no longer work.

about theoretical reduction and methodological autonomy, both of which invoke the relationship between scientific theories and their kinds. Worrying about functions and how they are brought about, on the other hand, is not going to make any ground in answering these worries about whether, say, psychology is autonomous from neuroscience. Figuring out whether the realization base has differences in properties relative to the individuative function gives us no clue whether the science of the realized kind is just another way of doing the science of the realizer kinds. So to the extent that multiple realization's history is rooted in these issues about reduction and autonomy, that is reason to reject a sense of multiple realization that is blind to these notions.

Just to be absolutely clear, let me elaborate on this point. The ultimate goal of Nagel-style reductionists was to demonstrate the unification of all scientific theories and, importantly, the reduction of all non-basic disciplines to some basic science (physics). While these goals may not inspire many philosophers of science or mind currently, they do constitute the historical milieu in which the identity theory was born. That view was, in effect, that mental state types, however those come to be individuated according to mature psychology (or whatever), are identical with (reducible to) brain state types, however those come to be individuated by mature neuroscience (or whatever).[12] It was noticed that issues about the reduction of theories carried with it ontological consequences about the relationship between the various entities picked out by each theory's descriptive vocabulary. The early identity theorists' hypothesis was primarily a response to ontologically extravagant dualist theories, but it was also plainly couched in terms

[12] Witness the ever present calls for caution with references to 'c-fiber stimulation'. That is transparently supposed to be a placeholder for the kind specified one day by the taxonomy of the science which covers brain state type individuation, which makes it even more evident that questions about multiple realizability of 'mentalistic' terms reflect a dispute about the relationship between two scientific classifications.

that make it an educated guess, an academic gamble, about the future of mentalistic and neuroscientific theories. Comparisons of the identity theory to other famous reductions (heat to high mean molecular kinetic energy, light to electromagnetic waves, etc.) are not intended to point out helpful analogies; identity theory is committed to reduction exactly like these we heard about in introductory philosophy of mind classes. Objections to the view on the grounds of multiple realization of mental state types thus inherit the assumptions about the identity in question being one of kinds articulated by different theories. Hence, ultimately, the other name which writers give to identity theory: *reductive* materialism. A type being multiply realized is accordingly a claim about how homogenous kinds in one science (typically referred to as the 'higher-level' science) can be instantiated in heterogeneous kinds from the perspective of another science (the 'lower-level' science). It is a claim about how one *theory* relates to another *theory*. So in order to evaluate properly whether some kind is genuinely multiply realized, all we have to do is make sure there is sameness of kind at the higher level accompanied by differences in kinds at the lower level—where, of course, 'differences' are cashed out entirely by appeal to the taxonomies of the theories which cover events of the lower, realizing level. If, to use the crude old saw, pain is realized in c-fiber firing in, say, humans, and the same pain is realized in d-fiber stimulation in, say, dogs, then pain is multiply realized. What makes us confident in asserting that we have distinct realization bases is that, from the perspective of the science which divides brain areas into c-fibers and d-fibers, the two bases are distinct brain kinds. That is how you figure out which are the genuinely different realizations.

How does my proposal relate to Shapiro's way of picking out differences in realizers? As I said, it seems to give us reason totally to ignore his constraint on multiple realization. It does not matter whether two different instantiations differ with respect to the properties responsible for realizing the functional kind in question. Whether or not two token realizations do, or do not, will not settle the question of whether those realizations are really multiple. I am sure there are some cases where the causally relevant properties of two token realizations will differ and so will the realizations from the perspective of the science under which they are subsumed. But these are only cases where Shapiro gets the right answer by accident. So two token realizations of some arbitrary psychological kind can be made of different stuff (neural nuclei in one case, silicone chips in the other) *and* realize the kind in causally different ways, and Shapiro and I will agree that we have two distinct realizations of that psychological kind. But if the kind is made out different stuff, but each realizes the kind in exactly the same functional way, Shapiro and I will disagree (see the discussion in the section below about silicone brains). He thinks what is important is the difference in the causes that bring about the functional kind; I think what matters is whether the scientific theory which governs the realizer kinds judges them to be distinct.[13]

Section 4.3: Criticisms of Shapiro's View

I hope that shows exactly how our views collide. I have tried to give some reasons I have in favor of my view of how to distinguish 'multiple' realizations. But even if my alternative is

[13] This issue about the different senses of multiple realizability runs much deeper than what I am letting on here. But it is, I think, beyond the scope of this essay to plumb those depths. If anyone is worried that my notion of multiple realization—steeped in the philosophy of science as it is—requires a more extensive defense than that provided, I think I can give it back without endangering anything in my overall argument. Whether you think non-scientific kinds/types/properties can be multiply realized, I take it you also have to think that, in the context of psychological kinds being multiply realized, psychology and neuroscience do all of the real work. That is, the individuation criteria for the realized thing and realizer thing are fixed by those sciences. As long as the more liberal, non-scientifically-oriented conception of multiple realizability admits that, I get everything I want (against Shapiro).

not acceptable, it is still worthwhile if I can show why we should not accept Shapiro's view either. At least doing that will remove the worry I have about using strong equivalence to tighten down sameness of psychology in the face of his view about what makes realizations distinct. To that end, let me now start on the negative project of providing a few criticisms of Shapiro's notion of 'multiple' realizations in order to further establish a case against his view.

First of all, the scope of his remarks about different realizations revolves entirely around *functional* kinds. He limits himself to these kinds, of course, because he simply assumes that the best chance for finding multiply realized kinds is by looking at functional kinds only. An obvious consequence, though, is that the view gives us no way of determining whether a *non*-functional kind is multiply realized. Let us just take Kim's example of jade. Jade is, or so the mineralogists tell us, actually two different mineralogical kinds: jadeite and nephrite. In other words, when you find a sample of jade, you have either got a piece of jadeite or a piece of nephrite. Now, it is not altogether clear if jade does actually specify a true kind—so much as it is just a commonly used word which has always unknowingly referred to two different kinds of minerals—so there is the problem here of whether jade is even a candidate for being multiply realized. But we can set aside that worry and still get across the import of the criticism. Blindly assuming for the moment that jade is a kind, how could Shapiro's view tell us whether it is multiply realized or not? Jade sure is not a functional kind; it is not at all typed by what it does, or functions it performs. So a fortiori there will be no differences in the causally relevant properties at the level of jadeite and nephrite (in how they bring about jade, the kind which we are supposing they realize). So this must mean that jade is not multiply realized. But is this not a bad way of determining that truth? Whether or not jade is multiply realized should not be trivially decided by the fact that it is not a functional kind.

As I admitted, the case of jade—while I hope illustrative—is definitely tendentious. But the same point will hold, I think, for cases where our intuitions and the sciences are clearer. Water, for instance, is a paradigmatic example of a kind which is not multiply realizable. It is very much uniquely realized in H_2O. But, again, water is no functional kind; it is not defined by anything that it does. So there will not be any differences in the properties that are responsible for carrying out the function that defines the kind that water is. In this respect, I suppose, water trivially is not multiply realized. But that is just the point I am driving at: the issue of whether water is multiply realized definitely is not a trivial outcome of its not being a functional kind. Water's not being multiply realized has everything to do with the empirical fact that it is always realized in H_2O. Given Shapiro's perspective about how to make this determination, however, we would be utterly blind to this all-important empirical fact. It is the blindness to this fact that I find objectionable. The case of the multiple realization of psychological processes seems to me precisely to parallel this case about water. Whether or not these processes are multiply realized is going to depend on the taxonomies of brain science, not on anything to do with the finer details of the nature of the processing which instantiate the psychological kind.

Hence it seems smarter to me hold that the limited scope of Shapiro's criterion ipso facto makes it unsuitable for determining which realizations are actually distinct—particularly when the issue we are ultimately trying to settle is the multiple realization of psychological kinds in brain state kinds, both of which are scientific kinds. In short, functional kinds are not the only kinds there are. So having a criterion for determining multiple realizations which only works for functional kinds does not sound at all like a fruitful way of determining which realizations are distinct or the same.

Next I want to scrutinize the argument Shapiro has for his criterion. This is somewhat difficult to do, unfortunately, because he is rather inexplicit about how that argument goes. A likely interpretation of it, however, looks something like this:

(A) 'the properties which are relevant for counting in the realization base are those relevant to bringing about the function which individuate the kind'

(B) therefore, 'what count as different realizations are those that bring about the function (which individuates the realized kind) in different ways'.

Shapiro starts with (A) because functionally analyzing some task provides an objective method for determining what the important properties are for the task. We move on to (B) because "it seems reasonable to suppose that judgments of sameness and difference between realizations might be relativized to those properties that make a difference to how a functional kind functions, that is, to how a functional kind actually manages to bring about the capacity (or capacities) that defines it as the functional kind that it is". (Shapiro 2004, pg. 52). In other words, Shapiro thinks that the former (particularly as a response to the question, 'Why is it that the color of the artifacts does not matter for realizing the corkscrew kind?') provides us with some kind of reason for thinking that genuinely different realizations are those which differ in those realizing properties. That the realizing properties must consist of the properties responsible for providing the function that types the object (as a corkscrew, mousetrap, or mind, etc.) is supposed to allow us access to the claim that *different* realizers are marked by differences in those functionally relevant properties. From here, after we have pinpointed the functionally-relevant properties, Shapiro seems to proceed abductively. With our principle in hand about how to identify distinct realizations, we should apply it to particular cases and see how things match up with our intuitions. In the case of corkscrews differing only in color, the principle seems to get things right. Likewise, in the case of the waiter's versus the winged corkscrews, Shapiro's suggestion seems to get things right. If his hypothesis is continually confirmed by the cases, then there exists some reason for accepting the hypothesis.

I think, of course, that the argument from (A) to (B) is illicit. But that is because I think (A) is false *and* the inference is invalid. Though I want to say quite a bit about why the move does not work, I will start with the former criticism about the falsehood of (A). We have good reason for doubting the truth of this claim, I take it, for the same reason I have just given about doubting the truth of Shapiro's conclusion—viz., that functional kinds are not the only kinds there are. If we want to put a constraint on what properties are allowed inclusion in realization bases, then we would be wise not to limit our remarks to functional kinds only. Jade, or water, or 5-hydroxytriptamine is not a kind which is individuated by anything that it does, though we do admit that these kinds are *realized* by other kinds at different levels (e.g., the chemical, or physical level). So clearly the realization bases for these respective kinds are not populated by any properties that contribute to these kinds carrying out the functions which are all important for classifying them. These kinds are not even classified by what they do. And since we still have cases of realized kinds, clearly it is not true that the properties which count in the realization bases are those which are relevant for realizing the function individuative of the realized kind.

The reason I have for thinking that the inference is invalid is a little bit trickier. There are a couple of ways I can think of to call it into question. The tempting response is to press Shapiro on the silicone chip brain example. That is, it seems to me that we have a case of two functional kinds (minds) realized exactly the same way from a functional standpoint, yet we would *not* classify the realizers as the same. In this way, we could try to challenge, as it were, the transposition of Shapiro's main claim/conclusion: if two realizations are in fact distinct, then it must be the case that there are some differences in the properties relevant for bringing about that function. Yet here seems to be a case where, by hypothesis, there are no differences in the properties that are responsible for bringing about the psychological kind, but we would want to say that the realizations are distinct. I think most philosophers of mind would agree, too; it seems very odd to claim that minds realized in brain meat, on the one hand, and silicone chips, on the other—even when they both perform exactly same function, from a suitably abstract perspective—do not count as *distinct* realizations. The claim is so stark, in fact, my first reading of Shapiro (2000, pg. 645, or 2004, pp. 57 – 58) only made me think he had provided a perfectly good *reductio* of his view! No way could the assumption of his criterion for distinguishing multiple realizations be true if a consequence of the assumption was *that* sort of claim. I know he wants to make it so that multiple realizations are not cheap, but is that not making them entirely too expensive?

I suppose what Shapiro would say is that he cannot buy this claim that the realizations are distinct. If he is seriously committed to his method for distinguishing distinct realizations, then it follows that there is no important difference from the brain and its silicone chip replacement. So in order to make a convincing case for the reductio, I need to come up with a reason for thinking that the contradicting premise is true (i.e., I have got to show that these two realizations are, in fact, distinct—contrary to the consequence of his principle). But this reason should seem pretty obvious. Since we are working in the context of the identity theory (and the relationship of the kinds of two scientific theories), the goal of multiple realization is to show how mental state types can be multiply realized in *brain state types*. But silicone chip configurations are patently *not* brain state types. These two types of things are accordingly distinct types of things, making them each multiple realizations of the individual's psychology. In turn, that fact contradicts a consequence of Shapiro's argument.

Notice also that my suggestion for how to identify distinct realizations also accords with the intuitions that people ordinarily have about this kind of case. Certainly the neuroscientists would classify as a distinct kind of thing anything that was purely composed of silicone chips, regardless of any functional isomorphism between particular brain kinds and that bag of chips, drawing the conclusion that such a state of affairs would constitute a multiple realization of a mind—exactly the intuition everyone seems to have. Though I do not usually mind flouting them, there is *something* to be said for the weight of intuitions against Shapiro about this example. Especially since Shapiro seems to put a lot of weight on how his method for distinguishing different realizations is true because it does the best job of explaining our intuitions about particular cases, this silicone chip case is a thorn in the side of the plausibility of his view. What he has going for his account is the ability to explain the intuitions we have about various cases (e.g., the corkscrews differing only in color). So to the extent that his view depends on the abductive move about how well his principle meshes with intuitions about particular cases, he has a problem with the silicone chip brain.

Next, I want to draw attention to other things that we type functionally and look to see whether differences in how the function is accomplished generally track our intuitions about whether the different individuals of the functionally individuated thing really are taken to be *different* tokens, or *different* realizations, of that type. I cannot claim much accuracy about these intuitions; again, here we are simply dealing with ambiguous cases. But I think they will be suggestive for evaluating Shapiro's argument. In general, one can take this move to indulge Shapiro's limiting the discussion to only functional kinds, but to reject that his conclusion about what constitute genuinely different realizations even follows on this assumption. That is, in the bigger picture, I do not like Shapiro's suggestion because it provides us with no advice on how to determine whether non-functional kinds are multiply realized. However, I also think, even if we allow him the starting point of worrying only about functional kinds, he is still not right to maintain that genuine distinct realizations are determined by differences in the properties that are relevant for instantiating the function that individuates the kind. Not only does this criterion not tell us anything informative in the *very* general case (the case involving non-functional kinds), but it also does not tell us anything informative in his particular case (the case that only involves functional kinds). In this way, I will try to provide an example that grants the truth of (A), then see whether the move to (B) works.

So this sort of line is the dissimilar to the silicone brain objection, because there I was assuming the functions were carried out by the same algorithm, but were intuitively typed as different, too (rather than as the same realization, as Shapiro's inference would seem to require). To that end, consider mechanical watches—not the modern electronically driven kinds, but watches that require springs to be wound and make that distinctive 'ticking' sound. I do not know very much at all about how they work, but I do know that they are not all wound in the exact same manner. Some watches like this have a key which is used to perform the winding, and others simply have a gear on the outside of the watch which can be turned directly by hand. But despite this small difference in the overall functional architecture of the watch, the right intuition seems to be to call them both mechanical, wind-up watches. They do not really seem to be different realizations of the kind 'mechanical watch.' There are of course slight differences in the functions by which they keep time, but that is exactly the point I want to emphasize. Shapiro moves from differences in functionally relevant properties to a difference in realization; but I think that inference is probably not even valid. There are differences in the functionally relevant properties of the mechanical watches, but they do not seem to force us into saying these are multiple realizations of that type of thing.

Accordingly, it is not clear at all that the argument Shapiro provides for his claim that different realizations are found by locating differences in the functionally relevant properties is compelling. Though typically functional differences in how some kind is instantiated by two individuals do track our intuitions about whether those individuals are genuinely *different* realizations of the kind in consideration (e.g., digital vs. analog watches, standard vs. automatic transmissions, or the different ways of adding I mentioned earlier), I do not believe that it is always the case. The mechanical watch example, I think, demonstrates that this inference is not always truth preserving.

I take this to be the same troublesome outcome of the attempt to make more precise the vague, ambiguous sense of multiple realization. The same remarks hold for corkscrews as much as they do for different watches. We have some rough sense of what count as distinct realizations of each, but nothing like a precise principle to which we can appeal. But that is not really a problem. In other cases—namely, cases involving scientific kinds—we have built in precision. Maybe that precision does not extend to the vague cases (there is no science of corkscrews, and it is not clear at all to what other science we are trying to reduce these non-existent corkscrew laws and generalizations), but there is also no pressing reason to rescue those cases from such vagueness. Rather than trying to force down a principle which only muddies the waters by rubbing against our intuitions about the vague cases and by contradicting independently constrained scientific cases, it is probably best to leave the vague cases vague and accept what accuracy we can get.

But regardless of how to adjudicate this issue about different senses of multiple realization, I think the foregoing reasons point to this overall conclusion: Shapiro is wrong to hold that genuinely *different* realizations are counted by appeal to differences in the functional properties that are relevant for instantiating the function individuative of the kind in question. Assuming those reasons are adequately persuasive, we can set aside Shapiro's position about what makes different realizations really different—i.e., for the purposes of evaluating multiple realization. Given the way I have preferred to establish the sameness of psychological state kinds (roughly by sameness of processing), it is particularly important that I can ignore him. As we go along, it will become even more important to have some compelling reasons for dismissing Shapiro's attack on multiple realization.

Section 4.4: Constraints on Brains

Before concluding, however, I want to discuss for a moment a topic that he often mentions in his work—in order to provide a further illustration of why it will be important to deny Shapiro's criterion of what makes multiple realizations really *multiple* realizations. Extending our discussion of Shapiro in this way, I hope, will enrich the reader's understanding of the importance of his criterion, particularly in comparison with my proposal about letting the sciences do the work of figuring out the genuinely distinct realizations. It will help to show more precisely what is at stake in thinking about how to determine which realizations are really distinct.

Shapiro talks at length about 'constraints on brains' and how these constraints bear on the plausibility of the multiple realization of psychological processes. In general, the business about constraints on brains is an attempt to show that brains are not, and maybe *could* not, be as multiply realized as we might suppose.[14] He asks us to imagine that the tape of life is replayed in such a way that evolution begins anew, but with the provision that organisms with exactly our psychological capacities must be the ultimate result. The interesting question is: would evolution have to create the same brains that we have now—with neurons and myelin, dopamine and glutamate, topographic maps and receptor cells, and all the rest? Or: would evolution hit upon a variety of ways to succeed with the task—ways that did not use neurons, neurotransmitter, and the like, but rather used other, hitherto unobserved means? Given that evolution is free to succeed with the task however possible, one who endorses the multiple realization of psychological kinds, it seems, would expect that there would be many different 'brains' that would evolve which are capable of instantiating those kinds. On the other hand, those who are hostile to the thesis of multiple realization—those who accept what Shapiro calls the 'mental constraint thesis'—would probably expect that a new round of evolution, one which is rigged ultimately to produce psychologies roughly like ours, would not produce any strange, unheard-of brains. Instead, the brains of this alternative evolution would be more or less just like ours. They would, that is, have to possess a lot of the same functional features that our brains do. Shapiro's answer, of course, is that brains capable of realizing just our psychology are much more constrained than is commonly appreciated. And it is this fact which tells against the plausibility of the multiple realization of psychological processes.

[14] I will lazily not bother at all to flesh out the modal part of this claim adequately. Shapiro, as far as I can tell, is never really clear about it either. And, as far as I can tell again, nothing much important hangs on getting super-clear about it—not for the purposes I have in starting with this thread, anyway. So in what follows I will imprecisely, and mostly vaguely, use phrases like 'inevitable', 'must', 'can only.' I ask the reader not to give them as much as attention as they usually deserve.

Just what are these all-important constraints then? Here is one: given our fairly keen visual abilities, it must be the case that we have very many, and at least a few small, photoreceptors on the retina.[15] In general, the idea is that we need some way of detecting (in this case) electromagnetic waves in the environment. After developing receptors for doing that job, it seems obvious that we can only provide as much information as we have information-processors. So if we want to be able to capture a lot of information—which will amplify, among other things, our discriminatory abilities—we will inevitably need very many photoreceptors. Think, for a comparison, of the resolution available in digital cameras. Roughly speaking—though it is not entirely accurate to say so—the more pixels which are employed, the higher picture quality (the higher resolution) obtained. In this case, the pixels are like the photoreceptors and the quality of the image mirrors the average human's visual acuity. More precisely, it also helps that the relative sizes of the photoreceptors are small. This will ensure that as much information as is possible will be captured about every nook and cranny of the visual scene. The result of very many, small photoreceptors—much like what is found in fovea of the human eye—is improved visual capacities. So the overall suggestion is that part of what is necessary for attaining our level of visual acuity (attaining our psychological ability to see) is having the number, size, and probably organization of photoreceptors found on the retina (or, at any rate, something that

[15] We must set aside the worry that receptors, transducers, and some common code in which the information is processed in the brain are also constraints on brains with our psychology—otherwise, you might not need *any* photoreceptors to increase visual discrimination, let alone many. We can set this concern aside, though, because the main point is simply to get across the idea of what a constraint would look like, not to do any of the heavy lifting required to show there really *are* such constraints. In any event, Shapiro does cite Ulinski (1984) as a source of argumentation for the necessity of transducers, et. al. The retina is also not a part of the brain, obviously, so this kind of constraint is technically not a constraint on *brains*. But, again, the main point of the example is to make sense of constraints, not to quibble much about the details.

performs the function of detecting electromagnetic radiation and making its presence or absence known to the organism).

For another example, which builds nicely on the foregoing (and which, if true, is a constraint on *brains*): consider the topographic organization of many of the sensory areas of the cortex, including the primary visual cortex (V1). Remarkably, in V1 (and elsewhere) the organization of the receptors on the retina is preserved faithfully in this area of the cortex. Cells in V1 which are responsible for encoding information of photoreceptors which are adjacent on the retina are adjacent in the cortex. The relative size of the cells in the cortex are all the same, so an interesting feature about our brains is that more cortex must be taken up to process the information coming from the areas in our bodies packed with more receptors. In V1, the fovea takes up most of the processing area, while the outer reaches of the retina take up less. In similar fashion, since our tongues, hands, and lips are so sensitive, these areas take up more of our somatosensory cortex. The famous sensory homunculi one sees in neuroscience textbooks provide an illustration of this kind of phenomenon. All of this is supposed to be an instance of one type of constraint providing another constraint further 'downstream' (in the overall processing of the information). Since the presence, organization, and size of (photo)receptor cells is a constraint, we can predict that the brains which figure in to the rest of the processing will also have a kind of organization which mirrors the distribution of (photo)receptors. So here is another proposed way in which nature has no choice but to produce brains with particular functional features (assuming the goal is to develop organisms with human psychological abilities).

As I have mentioned parenthetically as a possible worry, we could probably debate very seriously the extent to which these sorts of constraints are actually, as Shapiro calls them, ‘universal’ constraints. It is possible, for instance, that these are only (again, Shapiro’s words) ‘historical’ constraints—i.e., they are only traits an organism will have in virtue of having the genealogy it does, not in virtue of the fact that any organism, regardless of lineage, must have that trait in order to do the interesting things it can do. In other words, a topographic organization of the cells in the cortex may be a mere artifact of descent from creatures who processed visual information that way, not something that is really a universal constraint on any organism which possesses, say, the visual capabilities of humans. I suppose it is also a possibility that having, say, a topographically organized visual cortex is simply the optimal design (for visual processing). I do not know in what precise sense it would be ‘optimal’ (probably something to do with contributing to better overall fitness of the organism?), but it strikes me that a trait’s being optimal could allow it to be selected for without that trait’s subsequent prevalence having anything to do with genealogical history or even with its being a truly universal constraint. In this way, we could also come to question the ‘necessity’ of topographic organization in the cortex. Maybe such organization is optimal, but that hardly seems to imply a universal constraint on processing information. As with the adaptive value of any trait, part of what sets that value is the environment in which the organism finds itself. No trait is simply selected for, completely absent of ecology. If brain power and/or resources are abundant in some environment, efficiency of processing (assuming that is what is so important about the trait) will not be much of a concern.

But, though I think there could be a good case against Shapiro, I am not going to try at all to state it. The reason for this is that all this business about constraints, *assuming Shapiro’s criterion is no good*, is entirely orthogonal to debate about the plausibility of multiple realization. Of course Shapiro thinks that these constraints, if true, lessen the plausibility of multiple realization. If, in order to possess the psychology humans have, it is necessary to possess more or less exactly the kinds of brains we have (functionally speaking), it starts to sound as if the same kinds of mental states come combined with the same kinds of brain states. The problem only occurs, however, when we make the assumption that different realizations are found, loosely speaking, by differences in the ways in which the all-important functions are brought about. Shapiro is arguing that, owing to these universal constraints, there will be very limited ways in which one can perform the functions that realize, say, the visual ability of humans. According to his way of counting, there will not be any differences in the realization bases of human psychology. But if we reject the idea that different realizations are found in this way, which I have tried to do above, we open up the possibility that the processing is relatively fixed (for, e.g., human vision), though that processing is realized in different kinds of things. In other words, those working in artificial intelligence can copy down to the finest detail what they know about the way in which visual information is processed in the human brain, implement that processing in some object that is not composed of brain meat, and the question of the multiple realization of that process is still wide open.

It is in this last respect that I claim that the business about constraints is irrelevant to the question of multiple realization. Of course I think that, in order to establish that two creatures are psychological similar enough to provide sameness of psychological kind, algorithmic similarity of processing is very important. But, particularly, I do not think the question of whether this kind of processing (among other potential candidates) presents a real *constraint* on creatures with our psychology is going to settle the question of multiple realization. Maybe this kind of thing is a genuine, universal constraint, maybe it is not. I do not see why I should have to decide. So long as the processing that our psychologies apparently demand crosscut brain science taxonomy, I am not concerned with the way the evidence points. I only assert that, as a matter of fact, we should care about similarity in processing, since it provides us with the most rigorous check on sameness of psychological kinds that I know of.

Chapter 5: But Does Current Neuroscientific Research not Presume Identity Theory? A Reply to Bechtel and Mundale

The relevance of Bechtel and Mundale (1999) to my dissertation, in very general terms, is their rejection of the multiple realization of psychological state types. More precisely, they think that a proper examination of neuroscientific practices will show that psychological kinds are not, in fact, multiply realized and that cases which are typically offered in favor of these states being multiply realized do not constitute serious evidence. Since this paper carries much weight in the literature on multiple realization and reduction, it is appropriate for someone taking my position to have some kind of response to this general line. But that is not all that Bechtel and Mundale say which is significant from my view's perspective. Their examination of neuroscientific practice also reveals a conflict with what I have to say elsewhere about the how to type brain states. They agree in the abstract that a proper taxonomy of brain states is attainable merely by looking at the sciences. But they think that, after reviewing what it is that neuroscientists are up to, the brain state kinds they offer are different from the ones I do. So below I will spell out these two different criticisms and try to develop some potential responses.

We will start with the first, more general criticism. After explaining why it is that Bechtel and Mundale think that cognitive states are not multiply realized, I will be able more quickly to explain the second problem about how to type brain states. They begin by reviewing the typical, somewhat fanciful arguments offered in favor of multiple realization. The emergence of artificial intelligence in the 50s and 60s colored the view that mental activities are definable in terms of computational operations on symbolic representations. That is, instead of identifying mental processes with some sort of biological substrate, the prevailing winds favored a view of mental processes as the computations themselves. Thus the new important metaphysical claim became: what it is to be a mental process is to be a kind of computational process. Since computational processes are capable of implementation on a variety of hardware, proposals about artificial (non-biological) minds came to dominate. From there, given the significant difference between the bases in which human and artificial minds could be realized, talk about alien minds also became popular. Surprisingly, Bechtel and Mundale do not have much to say about either the metaphysical claim or any of the more speculative arguments from aliens and robots. These arguments, they say, rely only on the mere logical possibility of multiple realizability—something neither feels inclined to reject.[16]

Instead, they want to look at an important conclusion that is typically drawn from the empirical claim that minds are, in fact, multiply realized in existing biological systems. Their attack is twofold. First of all, some writers who are partial to the multiple realization argument often say things like: given that psychological processes are multiply realized in brain kinds, it is unlikely that physiological studies of the brain will offer much by way of theory construction in psychology. If psychological processes are instantiated in a potentially infinite number of physical substrates, then it must be that research focused purely on the realizer level will only provide irrelevant information from the realized level's perspective. It is a major goal of Bechtel's and Mundale's paper to perform a modus tollens on this conditional; they believe that, e.g., neuroscientific research has contributed greatly to the understanding of human psychology and they offer an extended example of just such a contribution to theories of visual processing. So far as multiple realization does imply that brain level research cannot inform psychology,

[16] Recall that I am in agreement with them. The argument I am trying to defend is that psychological kinds are more than just possibly multiply realized; they are in fact multiply realized.

then the success of neuroscience in helping to formulate information processing level theories shows the thesis that psychological states are multiply realized to be false.

The second line of attack, another sort of implication one might be inclined to draw from the multiple realization of psychological kinds, is that it is unlikely that comparative analyses of brains across species would even come close to providing a useful guide for psychological theory formation. If minds are multiply realized, then it seems intuitive, if not inevitable, that minds will at least be multiply realized across distantly related species. So using the brains of an elephant to provide hints about psychological processing in humans, on the assumption of multiple realization, would be next to pointless. But, as Bechtel and Mundale point out, it is patently true that most of the neuroscientific research (which has been so helpful with regard to understanding human psychology) is carried out on different species. So, again, if the truth of multiple realization makes it unlikely that this kind of comparative research would be fruitful, the very existence of that research is a knock against the plausibility of the view. Notice, of course, that this particular line of reasoning also depends on the empirical fact that minds are multiply realized, and not merely on the logical possibility of their being so realized.

Section 5.1: Lower-level Interventions are Not Irrelevant to Psychological Theories

About that first attack, then: are there examples of neuroscientific work that have had psychological import? Bechtel's and Mundale's main example is research into brain areas underlying visual processing. This work has a long history, extending even back to the 19th century. It was, in fact, around the late 1700s that Francesco Gennari first noticed the distinctive pattern of striation on the cortical surface in the back of the brain. But not until the work of

Henschen (1893) was it demonstrated that damage to various areas of what came to be known as the striate cortex resulted in various forms of blindness. For instance, he noticed that destruction of the entire half of the striate cortex resulted in blindness in the contralateral half of the visual field. Such information was not hugely important for guiding psychological theorizing, but it gave researchers some clues about where in the brain to search for visual processing.

At first, it was assumed by most that all visual processing was carried out in the striate cortex (the primary visual cortex). With the work of Cowey (1964) and Hubel and Wiesel (1965), particularly using single-cell recordings, it became accepted that there were multiple, connected topographic maps extending back into the various sulci of the occipital cortex. Thus areas V2, V3, and V4 were mapped. With these anatomical differentiations, it also became known that from V1 through V4 less and less specific information was processed. In the striate cortex, cells responded to highly specific features of highly specific areas of the visual field, while in V3 the features to which the cells responded became more numerous and the specificity of visual field expanded. For instance, in early areas of visual processing (in V1), individual cells will track features like the orientation of a bar of light for a very specific area of the visual field. In the later areas of the processing, cells respond preferentially to the presentation of whole objects, no matter where in the visual field they appear. Consequently, the standard view of the cognitive science of vision became that we build up from very local areas and specific features to a more comprehensive picture of the visual scene. On top of that, Zeki (1977) also indicated that cells in V4 were especially sensitive to the wavelength of the stimulus. Now cognitive scientists had very serious information to use in understanding the flow of information received by the retina, transduced, and then processed in the occipital cortex. We do not, for instance, merely 'see' the visual environment all at once. There is a rather specific construction of our conscious visual scene from very select features of the distal stimulus.

A major breakthrough occurred with the work of Ungerleider and Mishkin (1982), mentioned above in Chapter 3 (about typing psychological states). From work on lesions in monkeys, they were able to determine the existence of two major pathways extending from the striate cortex. One route proceeds ventrally into the back of the temporal cortex. Lesions along this pathway typically correlated with the loss of recognition of patterns or the recognition of previously presented objects. The other route from the striate cortex proceeded dorsally into the back of the parietal cortex. Lesions in these areas correlated in produced deficits in selecting response locations on the basis of visual landmarks. Hence the former pathway seems dedicated to figuring out 'what' an object is—by figuring out what kind of properties some object in the visual scene has, like shape, color, and texture. The latter pathway seems dedicated to figuring out 'where' an object is—by keeping track of spatial relations among objects. So cognitive scientists now had even better clues for understanding the processes by which our visual abilities are carried out. Noticing these effects of lesions in specific areas provided a great leap forward in the understanding of the nature of visual processing. Now researchers began to know much more about the breakdown of the flow of visual information at its early and later stages.

Complicating this rough picture, Felleman and Van Essen (1991) provided an exhaustive study which showed that Ungerleider's and Mishkin's two-pathway parsing (unsurprisingly) far understates the complexity of the processing. They posited no fewer than 32 distinct visual areas, each of which maintains on average connections to and from 10 of the other areas. A full 1/3 of the possible connections between the areas are actually realized. The basic result, without considering the staggering amount of details, is that the number and interconnections of the areas make the distinction between the ‘what’ and ‘where’ streams much more blurry. Such a simple decomposition is not far from the truth, of course, but the generalization does very little justice to the actual nuts and bolts of the flow of information. Presumably, more work in disentangling this new picture of how humans see will only help further refine the theories of processing cognitive scientists seek in modeling human vision.

So, in short, it is certain that research purely related to the brain (lesion studies, single-cell recordings, or even just tracing basic connectivity patterns) has guided our understanding of the visual information that is processed there. Clearly there are other methods psychologists employ to formulate psychological models of all sorts of cognitive tasks besides investigations into the finer points of brain physiology. That is clear from the fact that any serious psychological work has been accomplished well in advance of our understanding even the crudest facts about the brain. Psychology, of course, has not been pursued only since the late 1980s or so. I am sure the people of the Middle Ages even had some kind of rudimentary folk psychology long before they had a clue that the brain was important for anything much besides cooling the blood! But certainly it is also undeniable that research into brain activity has enhanced our understanding of the kind of information that is tracked and how that information is manipulated. To that extent, brain research is patently not irrelevant to the formation of psychological theories of (e.g.) vision.

For something short and easy to refer to, the argument is supposed to go something like the following:

P1: If psychological kinds are multiply realized, then research into the realizer level (e.g., neuroscience) is irrelevant to theory construction in psychology.

P2: (Owing to cases like neuroscientific work in vision) Research into the realizer level is not irrelevant to theory construction in psychology.

C1: So psychological kinds are not multiply realized.

Section 5.2: Why This Lack of Irrelevance is Irrelevant

So what now for multiple realization? By simple modus tollens: if the truth of multiple realization ensures that research into the brain must be irrelevant for psychological purposes, then multiple realization must not be true. So there are two options. I could try to deny what seems like an obvious fact that brain research has had psychological importance, thus rendering one of the premises false. But that sounds to me like a complete waste of time. Or I can try to show how Bechtel and Mundale are wrong to believe that multiple realization has such an entailment, thus rendering false the conditional premise on which the modus tollens turns. The second route, denying the first premise, strikes me as the most promising. In fact, I think this path is more than just promising. It strikes me as pretty clearly true that Bechtel and Mundale are wrong to think that multiple realization carries with it such a consequence. Since they have both pinned such a great deal of their case against multiple realization on this argument, it is worth taking a few pages to detail why they are mistaken.

First, we should ask just why anyone would come to believe this claim about what multiple realization entails. The discussion about the autonomy of the special sciences, psychology in particular, is most responsible for this confusion. Typically, those in favor of the autonomy of some science are denying that the way someone does this higher level science is just the way one does some other lower level science. So the methodology of psychology is autonomous from the methodology of neuroscience, for instance, when the practices of psychologists—how they collect evidence, how they measure it, or interpret it, or even just how conclusions are drawn—differ from those of the neuroscientists. A common remark to make in this context of the autonomy of psychology is to say how difficult it must be to recast generalizations and laws couched in neuroscientific terms in psychological cases. The move seems to be that genuine autonomy of two sciences would imply that progress in the one will have little effect in the success (or failure) of the other. Multiple realization causes trouble here because it is taken to show that psychology is not theoretically reducible to neuroscience, meaning, among other things, that the theoretical vocabulary of the former does not map in a one-to-one fashion onto the theoretical vocabulary of the latter, which should mean that the language of the one will at least not always be informative for the other (in terms of generating laws, or providing any important insight for theory construction). So advocates of the autonomy of the special sciences have historically latched onto the prospects of multiple realization.

As far as that goes, fans of autonomy should be happy to support multiple realization. Certainly if it turns out that their science of choice *is* reducible (to physics, neuroscience, or whatever)—if the argument from multiple realization crashes and they have no other alternatives—then the hopes of autonomy are completely dashed. Because if it turns out that the laws of the higher level science are essentially just different vocabulary for stating the laws of the lower level science, because the one reduces to the other, then it is not clear why anyone should bother, for instance, to conduct experiments in order to confirm higher level laws from that perspective only. Assuming for the sake of argument that psychology reduces to neuroscience, it would be like psychologists formulating laws, which were ultimately neuroscientific laws, without knowing anything about or trying to find out anything about the brain. It would be terrible methodology on the 'psychologist's' part. So the supporters of the autonomy should definitely be rooting for multiple realization.

But, that said, I do not see why multiple realization opening the door for autonomy also opens the door for the claim that brain research is irrelevant for guiding psychological investigations. Even if psychological processes are multiply realized in brain processes, the psychological processes are still token identical to the brain processes; they are still ultimately *realized* in those processes. To that extent, it seems unavoidable that a thoroughgoing understanding of the realizer level will definitely have something interesting to say about the realized process. Understanding the nature of the processing at the neuronal level, since the psychological processes are obviously instantiated by those neuronal processes, should at least tell you quite a great deal about that token psychological process. Given the identity of the two individuals, there must be at least a rough isomorphism between the processes such that figuring out the nature of the one automatically clues you in to the nature of the other. But notice that we can admit all of this without abandoning our commitment to the claim that psychological states are multiply realized. Psychological processes are, let us grant for the moment, multiply realized in brain processes. Still, *this* particular psychological process could be explained and understood purely in virtue of investigations into *that* particular brain in which the psychological process is realized—even though in some other kind of brain process, or in some machine process, the exact same psychological process is realized.

Let me provide an example. Though I do not think mousetraps are really multiply realized in the important sense I am trying to defend, they do provide rough, but intuitive, examples of the concept. And they will come in handy to drive home the current point. Imagine we present to someone a spring-loaded mousetrap who has never seen one before. I imagine that it would not take much time for that person, even if it were some child, to figure out the odds and ends about how all the bits of the mousetrap fit together in order to trap small animals. He would have a look at the spring, pull back the bar, then figure out about the part where the bait goes. He could probably even begin to say what would improve the performance of the token mousetrap—make the spring stronger, make sure the place for the bait is able to hold a rather large piece of cheese, etc. Similarly, he could say pretty quickly what would impair the functioning of the mousetrap—the bar is replaced by some lighter material that does not come down with lethal force, etc. All of these sorts of considerations, of course, are at the 'lower' level. They involve considerations about the actual bits of wood and metal out of which the mousetrap is constructed. But what is important is that we can easily figure out about the 'higher' level stuff (the overall function of the object) by, as it were, intervening at this lower level. But notice the really important point: *the person's ability to grasp the higher level functional picture from the lower-level considerations in no way shows that mousetraps are not multiply realized*. In a suitably loose sense of multiple realization, mousetraps are still very much multiply realized, despite the person's ingenuity in determining the higher level generalities from lower level considerations. His ability to use the configuration of wood and metal to figure out the overall function of the mousetrap does not rule out the possibility that different mousetraps work in very different ways. None of the others require a spring, and some probably do not even require any kind of bait. So, per impossibile, if there were some science of mousetraps, it could still be autonomous from whatever science dealt with the properties and functions of the realization level stuff. In short, since the function of trapping mice is instantiated in that spring-loaded mousetrap, of course one could use investigations at the realization level to guide an understanding of the function which is realized there. It is a simple example, but the point still remains: being able to inform our comprehension of a higher-level, functional process by understanding lower level comings and goings does no damage to the claim that those higher-level processes are multiply realized.

To be clear about this point, it is worth saying that I am not trying to show that it is possible to learn the *general* from the *particular*. The example I use is not supposed to say that one can learn about mousetraps in general simply by looking at the how the bits of this particular mousetrap works. Pretty clearly that is not very plausible. Learning about how the spring comes down and kills the mouse would offer no reason to think that one could catch mice with a really sticky floor. The same possibly holds true of the psychological case. It does not seem likely that knowledge gained by neurological interventions into the visual cortex of a macaque could count as a good guide to the workings of the primate vision in general. If primate vision is realized in different ways—particularly in functionally distinct ways—then knowledge of the nuts and bolts of the realization of vision in a macaque might be quite unhelpful for learning about primate vision in general.

Rather the point I am really driving at is that looking at the *lower level* need not be completely irrelevant from the perspective of the *higher level*, even though the two levels are independent. In order to figure out the more abstract, 'higher level' picture about the mousetrap, it is enough simply to look at how all of the individual bits of the mousetrap, the 'lower level' springs and bars, fit together. Though mousetraps can be designed in different ways, and though mousetraps with spring-loaded bars can be made of plastic or any other different kinds of metals, it is still possible to learn about the higher level facts of this particular device (viz., that a live mouse comes in to the trap, and dead mouse is the result) just by manipulating the more concrete features of it.

It works the same continually in the psychological case. In order to understand the psychological level process, neurological level interventions can be a good guide. For example, popular theories of recall of long term episodic memory maintain that, say, visualization of a witnessed event is important for accurate recall of visual details. When particularly fine-grained visual details are trying to be remembered, any kinds of visual distractions will disturb accurate recall by interfering with the visualization of the remembered event (hence the old advice 'close your eyes and try to remember'). One could find this fact out by interfering with the processing of the visual association cortex (e.g., by TMS). By intervening at this lower level and noticing that the higher level task of long term memory retrieval is disrupted, one could learn about the latter by looking at the former. But so far the multiple realization of the psychological process is left wide open. It is still entirely possible that a computer, or more simply a brain using different neurological state types, could realize the ability of humans to recall long term episodic memories. Nothing about the instruction of the higher level theory of recall by the lower level neurological analysis rules out the possibility of recall being multiply realized in distinct lower level processes.

To crystallize this point, let me introduce what I find to be helpful terminology.[17] On the one hand, we have the *autonomy thesis*, which says that the taxonomies, laws and research practices of one science do not reduce to those of a lower level science. Hence the specific methodology of psychology is independent of the methodology of neurosciences. How you *do* the latter need not at all be how you *do* the former, or vice versa. On the other hand, we have Bechtel's and Mundale's *irrelevance thesis*, which claims that the truth of multiple realization implies that "information about the brain is of little or no relevance to understanding

[17] I owe them to a conversation with Jack Lyons.

psychological processes" (Bechtel and Mundale, 1999, pg. 176). The point I have been driving at for the last few pages is this: a commitment to the autonomy thesis does not oblige a commitment to the irrelevance thesis. These two are quite distinct. A proponent of psychology's autonomy from lower level sciences can carry on all day long formulating authentic psychological laws in complete ignorance of brain function, chemistry, physics, biology, or whatever. But that is not at all the same thing as saying that the confirmation of those laws, or the construction of her psychological theories, is immune to criticism or influence from the brain functional, chemical, physical, or biological levels. The main argument that Bechtel and Mundale use against multiple realization founders on exactly this confusion.

There are some trouble-makers in the literature, I think, who are to be blamed for it. Fodor, for instance, is probably the most culpable.[18] As a philosopher of mind who is unusually hostile about the brain—particularly given his substantial empirical background (albeit in psychology)—he has said many things about the irrelevance of neuroscientific information to psychological research. My personal favorite is a diary entry of his in the London Review of Books where he wonders why people are always 'going on about the brain so.'[19] In this piece, he wonders why people are so enamored about the results of imaging research. Since so much money is sunk into machinery capable of creating PET and fMRI scans, he wonders what the scientific import of these results is. PET and fMRI scans are those pictures one sees of the brain with colors indicating the level of activation of brain regions while subjects perform some task. At bottom, as far as I can tell, he is only making the point that finding out *where* in the brain certain psychological functions are carried out (as opposed to, say, *how* they are carried out

[18] Bechtel and Mundale cite his (1975) on pg. 178 as someone who "explicitly contend[s] that, in fact, neural studies have not and will not enhance psychological understanding."
[19] London Review of Books, Volume 21 (19), pp 68 - 69.

there) makes no real psychological difference—or maybe even no serious neuroscientific difference either. He compares it to the importance of knowing where the carburetor is in the engine. Of course we would like to know that the carburetor aerates the fuel, that it is connected to the intake manifold, and so on, but we could easily get along fine as mechanics without knowing precisely *where* the carburetor is in the engine. Dumping all that money into imaging research, Fodor is suggesting, is like spending all your money on trying to figure out where the carburetor is. There is just so much more to be concerned with.

Now, when Fodor publishes things like this, the natural idea is to think he maintains the view that the brain is irrelevant for psychological taxonomies or research. But obviously these sorts of experiments are not totally irrelevant, and he even admits as much when he says that finding out that regions dedicated to thinking about teapots and taking naps, if they really turn out to be spatially adjacent, might just suggest an interesting psychological connection no one would have anticipated. He is joking to some extent, of course, but the point his point is clear: these images could conceivably have neuroscientific or psychological weight, but it is not obvious how much weight we are talking about, or even whether image researchers are merely taking pictures willy-nilly, without any ideas about what sort of evidence the images are supposed to provide. Either way, Fodor is not holding any staunch line about the irrelevance of the brain science to psychology. If anything, he is just rhetorically overstating the case for the irrelevance of certain *kinds* of brain research to the purposes of psychology. No one who thinks that psychology is autonomous from neuroscience thinks research into the brain can serve no psychological purpose—no physicalist, anyway.

Section 5.3: Why Cross-Species Analyses are Important for Psychology, and Why this Importance is Irrelevant to Multiple Realization

Enough about autonomy and irrelevance. The next fact worth discussing about cortical intervention driving psychological theory formation is that virtually all of the cortical intervention is carried out on non-human species. The typical animals used are monkeys (of all kinds of species), great apes, rats, and cats. Human subjects do make an appearance, but usually only because of disease, tumors, and Phineas Gage-like acts of God. And since their numbers are mercifully low, the studies which center on humans typically progress from a perilously small sample size that limits the reliability of the evidence they offer. Whatever the reasons for the prevalence of animal studies, they do raise the question—not just for Bechtel and Mundale—why other animal brains can serve as any kind of guide for *human* psychological taxonomy. I had said earlier that intervening in the token brain state could very much have implications for the token psychological state instantiated there. But why should cutting up a macaque's occipital cortex give us any information about the psychology of my visual abilities? Maybe the irrelevance thesis stuff is misguided, but surely *that* kind of fact makes advocates of multiple realization blush. To put the question differently: if multiple realization were true, would it not be rather likely that macaque brains would be totally different from ours (even if our visual capabilities are similar enough)? Maybe we would not expect that there would be significant enough intra-species differences (among humans, say), but if there is any strength to the claim of multiple realization, then it will surely be borne out by *inter*-species brain differences. At bottom, this seemed to be Putnam's main reason for holding that the identity theory constituted such an "ambitious hypothesis" (Putnam, 1967). But perhaps his reason is undermined by the methodological assumption of neuroscientists and psychologists that intervening in macaque brains provides a nice surrogate for intervening in human brains. This is another of the major arguments Bechtel and Mundale present against multiple realization.

There are a few responses, one major and the other rather minor. The major response is a common one: that the similarities in mammalian brains are the result of homologies (see S. Kim, 2002). Homologies are all the characteristics (not just physiological ones, though these characteristics are the ones we are presently concerned with) which are shared between two members of different species as a result of having a common ancestor. In other words, these characteristics are the result of the success some former organism had in its environment, which caused the trait(s) in question to be common to all its descendents. It is unquestionable that humans and, for instance, all species of monkeys and apes have a recent common ancestor. They would not belong to the same order of mammals if this were not so. Hence the claim on offer is that the similarity of their cortices (among other things, obviously) is a result of their lineage. This means that advocates of multiple realization have no reason to expect that profitable work on human psychology could only be accomplished on inspection of human brains. The conditional Bechtel and Mundale appeal to—if multiple realization were true, then cross-species psychological work would be impossible—is not true.

I do not know very much about the weight of scientific evidence in favor of the claim that macaque/human neurological similarities are just the result of a recent common ancestor, but I would bet very much that biologists could produce mountains evidence to carry enough weight for a compelling argument. And if the biologists could not, I am definitely sure that geneticists could. The complete mapping of the genomes of monkeys, apes, and humans have all been accomplished and shed light on our genetic similarity. That this could be so would explain the surprise we have in finding out that the vast majority of neuroscientific research is performed on members of other species. This fact does little damage to the prospects of multiple realization because we have an excellent reason for why the brain state kinds are not as radically different as we might have guessed. Because they are inherited traits from the same ancestor, because the common genes have been expressed as brains that help the organisms to survive and reproduce, we should *expect* that the macaque cortex is fairly similar to the human cortex. This is simply a result of the laws of genetics, so it counts as very scant evidence in favor of the claim that human-like psychological capacities are inevitably housed in human-like brains. Probably a great deal of behavioral data accumulated by working with macaques will be applicable to the behavior of humans. This is so because the physiological realizations of macaque and human psychology are so similar.[20]

The minor reason for challenging this line is that there is some small reason for thinking that certain psychological abilities are realized in radically different ways (in other species). I call this reason 'minor' because, to a large extent, it is the burden of the rest of the dissertation to make good on this claim. The current example is intended only as a brief illustration of the kind of evidence I will present later. Here I will not be laying all of my cards on the table just yet. I also call this response 'minor' because I am unsure of the empirical soundness of the evidence I am going to invoke (for this particular case). That evidence is the similar psychological abilities that we share with birds, though, it is known, they have radically different brains than we do. For instance, most species of birds have fairly good perceptual and spatial navigation abilities—just

[20] This reply might be weakened somewhat by the prevalence of research on rats, a species not as closely related to humans as the various species of monkeys are. They belong to an entirely different order of mammals. But I take it that the favoritism for rats in neuroscientific experiments has much more to do with how cheap they are to get, breed, and house, not because they will do just as well as monkeys as proxies for human neuroanatomy.

as we do, to a lesser or greater extent in comparison—though no birds even have a cortex, one of the indispensable ingredients in human perception and navigation. So, without delving too far into the details, there is certainly some reason to think that other species have radically different encephalic structures at the heart of their cognitive abilities. Whether their cognitive abilities are suitably similar to ours to ground a claim to multiple realization, I do not know and will not be doing any research to find out (not, at least, for this particular case).

To supplement this minor point a bit, notice also that we do not use any vastly different species from ours when performing important neurological research. When we are interested in studying the processing of auditory information in humans, we do not bother to see how bat brains work, for instance. This could be, as mentioned above, more a function of which species are more easily obtainable and more easy to work with, costs with research being what they are (probably rats are literally a dime a dozen). But I am sure researchers also know that, say, the brain of a shark is just far too different from ours to count as a useful guide for understanding human brain function, even though we might share with them many similar cognitive abilities. To my knowledge, there are no studies comparing the similarity of any organism's brain that is not in the mammalian class with human brains, much less outside of the vertebrate phylum. I take it this has everything to do with phylogenetic relationships—and almost nothing to do with the correct metaphysics of the mind.

So, particularly given the point about the homologies that hold between humans and monkeys, I think we have enough information to vitiate the argument Bechtel and Mundale have about the prominence of cross-species research. If monkeys and humans did not share a fairly recent common ancestor and these similar brain structures and psychological capabilities evolved independently, then they would really have a plausible line. As things stand, there is a perfectly good explanation for why there has been such fruitful work on human psychology by studying macaque brains: we share a common ancestor from whom we inherited a lot of the same characteristics, including characteristics that shaped our brains. So opening up the heads of monkeys pays dividends for human psychology to a large degree because we have the same brains (the same realizers)—though this point is irrelevant to the debate about multiple realization owing to that common lineage. As we will see in the last chapter, there are actually differences in the brains of humans and monkeys despite psychological similarities.

<u>Section 5.4: Bechtel and Mundale on Brain State Types</u>

So enough about neurological interventions in psychology, and cross-species brain comparisons. Let us now move on to the next point that Bechtel and Mundale draw attention to in order to discredit the plausibility of multiple realization: the way in which neuroscientists map brain states. As I have said above in other contexts, for the purposes of gathering evidence for or against multiple realization, we have to settle on some way of individuating the psychological and brain state kinds. If we are not straight on the taxonomies beforehand, then no appeal to empirical cases will do any work. The interesting claim that Bechtel and Mundale make is that, in fact, the notion of a 'brain state' is a "philosopher's fiction" (pg. 177). That is, they think that the way Putnam characterizes brain states as, roughly, physical-chemical states of the brain is nowhere near the characterization of brain states with which neuroscientists work. That seems to be a far too fine-grained construal. Instead, neuroscientists are more concerned with activation of particular parts of the brain. But, what is most important, they also incorporate an appeal to function in interpreting these activations. So, in short, working neuroscientists use functional criteria in their brain taxonomy. What makes something a visual cortex, for instance, is that it is a part of the brain dedicated to processing visual information. Location, cytoarchitectural profile, and other features are also relevant to carving things up at the joints, of course, but what psychological function some piece of brain carries out is also kind-individuating.

Why is this point significant? Why should Bechtel and Mundale use it to argue against multiple realization? Well, as I have already made clear in previous chapters, if it is true that brain regions are individuated by psychological function, then the prospects for a psychological process(/function) being instantiated in truly distinct brain state kinds are dismal. In fact, it would not even be possible. Once the relevant psychological process had been identified, assuming that would provide an individuating criterion for the brain state, then whatever brain bit (call it B1) is correlated with that process is automatically marked as a particular kind. Take some other piece of cortex (call it B2): no matter whatever differences (or similarities) B2 has with B1 (at any level of analysis), so long as it was correlated with the same psychological function, B2 would be classified as the same kind of brain state as B1. So on the assumption that function is the essential element in brain cartography, as long as two creatures share the same psychological functions, then they must automatically share the same kinds of brain states. Multiple realization, in a sense, would be left dead in the water.[21] There would never be a case of same psychological function instantiated in different brain state kinds because sameness of the first guarantees sameness of the second. *That* is the significance of neuroscientists using function as a tool for and a criterion of brain state individuation.

[21] I explain below what I mean by 'in a sense'. As I have already expressed in the chapter about typing brain states with reference to a purely brain science taxonomy, I think such a mixing of psychology and neuroscience destroys the possibility of multiple realizability, but at the expense of the coherence of the identity theory itself.

Now, here in particular I really need to be precise about the way I am reading their argument. Bechtel and Mundale are not very clear about it, so I will present two different versions for how it could go.[22] First is the weaker version—called 'weaker' because of the strength of the second premise:

P1: If psychological states are multiply realized, then neurological classifications are not (even partially) psychologically/functionally-guided.

P2: Neurological classifications are *partially* psychologically/functionally-guided.

C1: So psychological states are not multiply realized.

Second is the stronger version. It would go like this:

P1: If psychological states are multiply realized, then neurological classifications are not (even partially) psychologically/functionally-guided.

P2*: Neurological classifications are *only* psychologically/functionally-guided.

C1: Psychological states are not multiply realized.

The real difference comes in how one is supposed to read the second premise. Is it that neurological kinds are only typed with reference to psychological functions? Or is it only that psychology is one of perhaps many individuative criteria? I think that both arguments are bad, but what makes them bad will be different for the how you read the second premise. As far as I can tell, Bechtel and Mundale vacillate between both the starred and non-starred readings of P2.

Section 5.5: Criticizing the Idea that Brain State Types are Functionally Individuated

[22] Later I offer yet another version to cover some other bases. It should be said at this point that Aizawa (2009) reads Bechtel and Mundale basically the same way I do here—in particular, he seems to believe that Bechtel and Mundale hold P1.

What sorts of responses are there to these arguments? I have three. First, it is simply an exaggeration to emphasize the role of functional considerations within neuroscientific taxonomy. It is true that if neuroscientific kinds are individuated purely by reference to psychological function, then multiple realization does not have a chance. But neuroscientists just plainly do *not* individuate types of brain states wholly by reference to psychological function. That is, P2* is utterly hopeless. Why should we think standard neuroscientific practice does not lay such emphasis on psychological function for individuating its kinds? There are numerous reasons. Many of the common brain areas referred to in neuroscience, for instance, are individuated without any reference to function: the lateral geniculate nucleus, the orbitofrontal cortex, the dorsolateral prefrontal cortex, or the substantia nigra. All of these areas of the brain are clearly identified by the location you would expect to find them or other various ways in which the areas appear—not by any function they perform. Bechtel and Mundale citing Brodmann is also very confusing. They provided quotations from Brodmann explaining that he was ultimately after a typing of brain states by localizing psychological function, but certainly he did not come up with the famous Brodmann maps by reference to any psychological functions. It is clear, as Bechtel and Mundale admit, that his individuating criteria were purely cytoarchitectural. Besides this curious allusion to Brodmann, it is also clear that modern neuroscientists still respect Brodmann's taxonomy, a fact which further drives home the point that neuroscientific brain-mapping is not *only* functionally guided. So the stronger, starred argument has no chance.

Secondly, the falsity involved in P2* is of a special kind. It is just flatly false, but one almost has the temptation to say that it *could not* be true—at least not in the context of the dialectic between the identity theory and multiple realization. Using this premise in an argument against multiple realization, if it were true, would be like some sort of category mistake. It totally overlooks the point I have already made that an evaluation of multiple realization and the identity theory must be made by sciences that are at least taxonomically distinct.

To explain, even if it is true that neuroscientists do use only functional considerations to guide their brain taxonomy, I do not think they are allowed to say multiple realization is false. Remember how I have been construing multiple realization: as a claim about how two scientific taxonomies relate to one another, where a single 'higher' level kind is instantiated in many distinct 'lower' level kinds. In order to evaluate properly whether some kind is multiply realized or not, we need to compare it with kinds from another science. But assuming that everything Bechtel and Mundale have said about brain taxonomy is true, if our question is whether some psychological process is instantiated in distinct brain process, then we have no way of coming to a definite answer. In this case, the taxonomy of one science (psychology) is plainly guiding the taxonomy of another (brain science). But, if anything, that just means we are dealing with two sciences which are not distinct—where the kinds of one are crammed in to the kinds of the other. But claims about multiple realization are not evaluated from within a single level like this. It makes no sense to wonder whether, say, chemical kinds are multiply realized in other chemical kinds. In a sense the answer would be 'no', but that does not mean that chemical kinds are not multiply realized (in physical kinds, e.g.). Claims about multiple realization are evaluated from a multi-level perspective, as an evaluation of how the one level relates to the other. From this point about functionally-guided brain state typing, it seems one can only really draw the conclusion that multiple realization would be simply undefined for the relationship between the two sciences. I apologize to the reader for making this same point I have repeatedly made up to now. Its requiring repetition in different contexts, I hope, shows how important—and yet completely overlooked—it is.

The third response I have focuses more on the first, weaker argument Bechtel and Mundale could be advancing (though what I say could apply to both arguments). I think the second premise in the weaker argument comes close to being false, as well, but I will accept it for now for charity's sake. That is, I think psychological/functional factors do not make much of a difference at all to neuroscientists. They do care about the structure of the brain and in particular how various mechanisms bring about psychological capacities (assuming 'structure' and 'mechanism' count as functional considerations), but that is a far cry from claiming that the brain state type it belongs to is fixed by the psychological service it performs. Neuroscientists, for example, are interested in how the hippocampus functions. Anything that does what normal hippocampi do would probably be typed as one—assuming it was made of the same kinds of cells, etc. But if it turned out that the mechanisms there carried out auditory processes, I find it hard to believe it would be classified as anything but a hippocampus. It would be a noteworthy hippocampus, but not less of one for having nothing to do with long term memory. But let the second premise be true, anyway. Even if it is, it will not help because the first premise is not true. The truth of the multiple realization of psychological kinds does not imply that psychology/function can have absolutely *no* classificatory weight.

To see why, I appeal again to the helpful cross-wired ferrets. To keep the details brief (because I have already discussed this case above), Mriganka Sur and his colleagues introduced a particular kind of lesion in the left hemispheres of the brains of neonate ferrets, which had the result of projecting the optic nerve to the MGN (rather than its customary stopping point of the LGN). This thalamic area, usually associated with hearing in mammals, in turn projected further to the primary auditory cortex. The surprising result is that, when the researchers tested the right visual field of the ferrets when they were adults (the optic nerves project to the contralateral half of the brain—the right eye's optic nerve to the left hemisphere of the brain and vice versa), some visual acuity was detected. Somehow, the auditory cortex, an entirely separate area of the cortex from the visual cortex, was analyzing visual information from the right retina and doing some processing. The acuity was, perhaps predictably, poorer than that of the left visual field (which maintained its typical projections from the retina to the LGN and on to primary visual cortex).

But, despite the lesser acuity, it is clear that the auditory cortex of the rewired adult ferrets was supporting the visual psychological processes. This cortex even came to resemble the early processing areas of the primary visual cortex, looking surprisingly little like the primary auditory cortex of the control ferrets! For instance, the topographic maps of the retina appeared and the orientation-sensitive columns which Hubel and Wiesel discovered in V1 also formed. The important point for my purposes, obviously, is that no one came to call the left auditory cortex of the rewired ferrets the 'visual cortex'. If it is true that functional considerations have an important role to play in brain taxonomy, then the rewired hemisphere ought to have brought with it a unique taxonomy of its parts. From a psychological, connectivity, and mechanistic standpoint (all functional considerations), the bit of cortex neuroscientists would usually call the 'auditory cortex' should have been classified as the 'visual cortex'. But of course it was not. In fact, the very fascination the experiment evokes *depends* on functional facts not having a very large role to play in taxonomic concerns. If they did, no one would find it interesting at all that this bit of brain over here (not in the usual location of the cortex compared to other ferrets), when connected in the right ways with upstream stuff, and when being correlated with the right functional processes, could do the things observed in the rewired adult ferrets. That would just be another case of the visual cortex doing its thing, albeit in a different spatial location than is typically seen in ferrets.

Notice this further curious fact, which sheds some slight doubt on even the brain taxonomy I proposed: even after the bit of cortex which was receiving the strange input came to *appear* like the primary visual cortex of normal ferrets, it was still referred to as the auditory cortex. That is somewhat strange because it seems to weaken the idea that even cytoarchitectural features have that much to do with the actual taxonomy of the brain, too. I think, in the end, that this line does not carry much weight against my way of counting the brain state types. One would have to show that, for instance, the layering of that bit of the cortex had substantially changed, or that the cell densities across the layers were now different, or that now there were different kinds of cells present. But, again, I will wait until I treat this experiment more fully to deal with this point. To be brief for now, I actually think it is precisely the cytoarchitectural profile of the auditory cortex of these ferrets which accounts for their decreased visual acuity (compared to their left visual field). In other words, we can still classify this area in the rewired ferrets as the auditory cortex because the persisting cytoarchitectural profile it has (which, according to Brodmann, is what classifies it as the kind of cortex it is) weakens its ability to perform the processing at a level that matches that of the right visual cortex. That there is such a drop-off in acuity is what stands as a marker that the area is appropriately classified differently (as the auditory cortex), even though it shares a lot of the same functional properties as the right visual cortex.

So, to conclude this objection, I think the case of these rewired ferrets probably does show us something about how brain parts are typed. In particular, I think it shows that Bechtel and Mundale are wrong to emphasize the extent to which functional considerations play a taxonomic role (assuming they are leaning towards something like P2*) and I think they are definitely wrong to suggest that "[f]or multiple realizability to be a serious option, brain taxonomy would have to be carried out . . . independently of psychological function" (ibid., pg. 177). For not only is it apparent that psychological function (connectivity, etc.) is not the only crucial factor for mapping the brain, it also seems like evidence for multiple realization can be obtained in spite these sorts of criteria playing a cartographic role. That is, I think—and will argue later on—that these cases of rewired ferrets do give us some reason for believing psychological processes to be multiply realized in brain processes. This may strike the reader as confusing since I did admit that the rewired ferrets have lesser visual acuity in the right visual field. Recall that I suggest we adopt a fairly strict notion of sameness of psychological process, so it seems implausible that these rewired creatures do share the same psychological processes as other ferrets. But I will deal with these worries later. For now, assuming I can persuade the reader about what sort of evidence for multiple realization these ferrets provide, it will be enough to show that Bechtel and Mundale are wrong about the significance of psychologically guided brain taxonomy.

I should admit, before concluding, that the strong version of their argument is perhaps implausibly too strong. Though they are not very clear in their paper exactly which individuating factor, or factors, they think are important, it is probably a straw man to construe their view so strongly. The reason for the strong construal, however, is that the inference to the impossibility of multiple realization only follows from this strong reading. This is the main reason why I considered two versions of what their argument could be. It is only when neurological types are picked out with reference to what psychological function they execute that psychological kinds will not ever be instantiated in any more than one neurological kind—viz, the neurological kind which is typed by reference to the psychological kind in question. If their view is supposed to be, on the other hand, that psychological function is just one of many individuating factors, then the threat to multiple realization is severely diminished, rendering their paper much less interesting. For other factors, like cytoarchitectural profile or mere location, multiple realization is left open as a possibility. That is, the weaker reading makes it possible that two tokens of the same psychological type are instantiated in neural state types which are distinct (because they are mainly classified by, say, cytoarchitectural profile). The goal of the final chapter of the dissertation is to make good on this possibility of the weaker reading: to find examples of the same psychological kinds being realized in neurological state types which are distinct because they are (mainly, if not entirely) classified by location and/or cytoarchitecture.

But one might press on behalf of Bechtel and Mundale that even the weaker reading is a bit too strong. In particular, one might think that they do not really believe P1, that the truth of multiple realization requires that neuroscientific states are not functionally-individuated. Again, they never do get precise on how their argument is supposed to go, though they do provide an entire section where they discuss the importance of functional criteria for brain mapping. And there are quotes in the text which support this reading of the premise. First, already mentioned, they say (pg. 177) that "[f]or multiple realization to be a serious option, brain taxonomy would have to be carried out . . . independently of psychological function . . ."; second, they say (pp. 204 – 205): "Moreover, as we have emphasized, it is the functional characterization (the contribution to behavior) that guides the identification of brain areas. A thoroughgoing functionalism uses functional criteria to identify both psychological states and brain states . . ." So there is some reason for construing the argument from section three of their paper in the way I have.

But maybe the best way to view their argument is negative. Rather than saying what multiple realization requires, perhaps they are merely undermining evidence that supporters of multiple realization have used to support their position. In particular, it might be thought that they are arguing somewhat as follows:

P1. If multiple realization is true, then psychological kinds are functionally individuated and neuroscientific kinds are not functionally individuated.

P2. It is not true that neuroscientific kinds are not functionally individuated.

C1. Therefore multiple realization is not true.

The typical supporter of multiple realization champions the first premise, all the while assuming that neuroscientific kinds really are individuated in ways that have nothing to do with psychological function. The contribution of Bechtel and Mundale is thus to point out that psychological function is at least one of the individuating factors. This is a happy way of reading their overall argument because P1 really does have some sort of life to it. Multiple realization does, it seems, require that the individuating factors for the relevant kinds be different. If, on the other hand, they are the same, as I have repeated many times, then multiple realization is definitely not true. Also, Bechtel and Mundale have a lot of historical evidence that supports P2 (some of which I talked about in Chapter 3). The great majority of their paper just is the description and discussion of this evidence about what psychological factors have done for brain mapping. So let us be charitable and assume briefly that, contrary to whatever you might read in the paper, this is one of their main arguments. What is wrong with it?

Yet again I think things turn on how we are supposed to understand the extent to which neuroscientific kinds are functionally determined. In particular, when the supporter of multiple realization puts forward P1 and claims that 'neuroscientific kinds are not functionally individuated', does she mean they are not *fully* functionally individuated or not *even partially* functionally individuated? As I have been at pains to say, the different readings here make a big difference. The truth of the premise turns on how we are supposed to understand it. Notice the resulting ambiguity in P2 between reading it as 'neuroscientific kinds are *fully* functionally individuated' (and hence it is not true that neuroscientific kinds are not functionally individuated) versus reading it as 'neuroscientific kinds are *partially* functionally individuated' (and hence it is not true that neuroscientific kinds are not functionally individuated). Once we get clear on how P1 goes, the only way to derive the falsity of the antecedent of P1 is to assert either one of the new readings of P2. But once we do I believe we find that either argument fails.

If we read P1 as 'If multiple realization is true, then . . . neuroscientific kinds are not *fully* functionally individuated', then I think the premise is true. As I have stressed, it is obvious that psychological and neuroscientific kinds being individuated along the exact same criteria spells the end of multiple realization (along with the identity theory, for that matter). Multiple realization's truth requires this consequent. However, in order to draw the damning conclusion, Bechtel and Mundale would need it to be the truth that neuroscientific kinds are *fully* functionally individuated. But, as I have discussed, that is clearly not true. How about reading P1 the other way: 'If multiple realization is true, then . . . neuroscientific kinds are not even partially functionally individuated'? I think the consequent-denying premise would be plausible (that neuroscientific kinds are partially functionally individuated), but, in this case, I do not think that P1 would be true. I have tried to show why multiple realization is compatible with neuroscientific kinds being at least partially functionally individuated, so there would be no reason for a proponent of multiple realization to assert with this construal of P1 that multiple realization requires functional criteria to play absolutely no type individuating role in brain taxonomy. So either way you go the new argument is not going to work. In one case the conditional in P1 is true, while P2 is false, and in the other case P2 is plausibly true while its P1 conditional is false.

So that is enough with the argument from brain taxonomy. To summarize the chapter for a moment, they have two main attacks: first, they point to the fact that fine-grained intervention at the brain level has provided psychologists with clues about which psychological theories to prefer; and, second, they show that neuroscientists use psychological considerations to guide their own taxonomies. Both facts, according to Bechtel and Mundale, are inconsistent with the truth of multiple realization. I have tried to show, using a few different arguments, why they are mistaken to believe there is such an inconsistency.

Chapter 6: But Do Cases of Neural Plasticity not Constitute Bad Evidence for Multiple Realization? A Reply to Polger

Another writer who has had influence within the literature I am considering is Thomas Polger. His attacks on the multiple realization arguments share a lot in common with Shapiro, but there are two themes, for lack of a better word, which appear in his works which are distinctly enough his own to deserve attention. Since it was upon reading his (2002) that I first started to think about all the issues I have been discussing, I want to take a few pages to examine these two general lines of criticism. On the one hand, he presents an interesting critical analysis of the argument I am defending—so that responding to his work, particularly given its prevalence in the multiple realization literature, is not out of line with the overall intent of this dissertation. But also, on the other hand, reviewing these two themes that he presses against (particularly) Putnam and Fodor will hopefully provide a useful introduction for the empirical evidence I want to invoke in the last chapter in favor of multiple realization. In particular, Polger is very upfront about what little evidence for multiple realization he thinks cases of *neural plasticity* amount to. Since I want to appeal to just such cases as empirical examples of psychological state types being multiply realized, it will be important to confront this theme in Polger's works.

Section 6.1: Begging the Question against the Identity Theorist

Before getting to the plasticity stuff, though, I want to devote a few lines to the other major theme. I call it a theme, though perhaps it is better named an argumentative strategy that Polger employs repeatedly in his works (see his 2002, 2004, and 2009)—viz., that many arguments against identity theory simply *beg the question*. When reading his papers, one cannot help but notice the frequent use he makes of pinning this fallacy on certain hypothetical moves one could make, or sometimes just on arguments that other authors have in fact made. I want to zero in on two instances of this kind of attack in his works: in his (2002) and his (2009), where I think this sort of defense of identity theory is not compelling. That Polger appears to favor this kind of attack on arguments against identity theory strikes me as a good reason for taking the time to point out these individual cases where he comes up short. Anything that will help to guard against confusions in the literature is worth laying out, I think. Also, these supposedly question begging arguments I will consider are also quite close to the arguments I favor (and will try to substantiate) against the identity theory, so, while not necessarily composing genuine obstacles to my view, it is worth pausing over this theme in Polger's work and trying to sort these arguments out.

Let me be clear from the outset, however: I do not intend to say that whenever Polger uses this response that it carries no weight. And I certainly do not mean to cast doubts about whether this informal fallacy really does constitute a fallacy. Of course there are examples of arguments that 'beg the question' in a way that is plainly illicit and contrary to sound reasoning. Rather, this informal fallacy is the trickiest of the bunch and deserves much scrutiny before it is brandished within the dialectic of a philosophical dialogue. When we look more closely at the arguments Polger labels as question-begging, I think we will find he has no good reason to do so.

Let me step back for a moment, though, and try to make sense of what I have just said by talking about begging the question in general. As anyone who has ever tried to teach the informal fallacies to students in an introductory logic course knows, begging the question is definitely the trickiest of the informal bunch. The main problem is that there are a few different ‘kinds’ of arguments, so to say, that beg the question. The textbooks will tell you that all of them have something importantly in common: that all such arguments that commit the fallacy create the *illusion* that inadequate premises (inadequate because of some sort of presumption) sufficiently support the conclusion. But each different ‘kind’ does so in dissimilar ways. For instance, there are arguments which plainly reason in a circle, where the conclusion and the premise, though identical statements (not just identical propositions), are separated by other premises which hopefully camouflage the repetition of the conclusion in one of the premises. Such an argument is obviously valid (just as ‘A, therefore A’ is), but counts as fallacious because of the attempt to paper over the triviality of the inference by introducing other distracting premises. Another ‘kind’ of argument that begs the question is one where a shaky, though key, premise is left unstated in order to avoid making it obvious that the soundness of the argument is clearly compromised by it. There are also cases where one of the premises of the argument, though a distinct statement from the conclusion, simply presumes the truth of the conclusion. In this case, accepting the dubious premise at all already gets you the conclusion, so that it would be redundant to use the premise as support for the conclusion. Finally, though you will not find anything exactly like this in textbooks, arguments are popularly said to beg the question when they run afoul of certain dialectical niceties. In particular, if an opponent has no more reason to accept a premise than she antecedently had to accept the conclusion, that premise is said to beg the question. That is, if someone occupies a particular position, it is dialectically objectionable to offer a premise in support of a conclusion that denies her view if she has no more reason to accept the premise than she does the denial of her view. It is settled within the dialectic what the issue for debate is; if a premise does not offer any more reason for belief than the conclusion which bears on the issue, then clearly that premise can do nothing to move the dialectic forward.

Hence this is a way of taking for granted (i.e., begging) exactly the question which the dialectic is about.

Now, I am fairly certain that the basis for Polger's objections is from this last interpretation of how arguments can beg the question. I think he is attempting to make the case that the premises offered in support of the claim that the identity theory is false are such that he has no more reason to accept them than the denial of his view. That is, in the same way a moral consequentialist would find the claim 'Consequences are simply irrelevant from the point of view of moral evaluation of actions' dialectically objectionable, so, too, Polger is going to say that certain premises used to show that identity theory is false are illicit. Since these premises—the ones I am concerned about—invoke multiple realization, it is worthwhile to discuss them, if for no other reason than to clear the way for my case against identity theory without having to worry about Polger's line.

So where is it that Polger wrongly accuses others of begging the question, where we can unpack this whole issue? I will mention two different lines from separate papers, just to give the reader a taste of this move he likes so much, then provide a response for them both all at once. The first example occurs in his (2009) paper (pg. 461) in a discussion of Fodor (1974). He quotes Fodor as saying "it is increasingly likely that there are nomologically possible systems other than organisms (viz., automata) which satisfy the kind predicates of psychology but which satisfy no neural predicates at all." This claim is supposed to support the idea that nomological coextensions of psychological and neurological kinds do not exist, thus undermining what Polger calls the 'brain state theory.'[23] Polger, in turn, rejects this argument simply by saying that "this reasoning we have already seen to be question begging." For now I will skip over explaining

[23] On pg. 458 of his (2009), Polger makes it clear that 'brain state theory' is the identity theory: " . . . as is the belief that this fact decisively refutes the identity (brain state) theory."

why he responds this way and move on to the second example. His discussion of the question-begging nature of the other example is more explicit and will apply equally, as far as I can tell, to this line that Fodor takes.

The second example of Polger playing the question-begging card is in his (2002). Here he divides up different varieties of multiple realization claims that one might use to assail the identity theory. Of the four he mentions (he calls them Weak-MR, SETI-MR, Standard-MR, and Radical-MR), I want to focus particularly on the response he makes to the intuition about Standard-MR. He construes this intuition as: "Systems of indefinitely (perhaps infinitely) many physical compositions can be conscious." To begin with, we have to set aside the fact that I am not concerned with conscious states, and that I also have been framing the thesis of multiple realization (by which we argue against identity theory, anyway) not in terms of physical state types but brain state types. I care mostly about psychological kinds being multiply realized in brain science kinds, not whether conscious state kinds (whatever those are) are multiply realized in physical kinds. We can safely forget these issues, though, because I think Polger would make the same criticism of either way of stating the intuition. Also—as should be obvious—I do think that psychological kinds being realized in kinds other than brain science kinds (any other physical kinds would do) is clearly evidence against the identity theory. So I do not object at all to the emphasis on indefinitely many physical compositions satisfying psychological descriptions.

Polger's begging-the-question claim about using Standard-MR against the identity theory is much more worked out, so I will explain his response in depth. He begins his response to this intuition by maintaining that "if functionalism is correct then Standard-MR is likely", though of course "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". (Polger, 2002, pg. 147) Because "if one must invoke a philosophy of mind in order to justify limiting realizers, then to whatever extent the justification is incompatible with Identity Theory it is question begging". (ibid, pg. 149) So the main idea behind this charge of question-begging-ness appears to be that, in order to generate any presumptive force against the identity theory, one must not appeal to intuitions which depend on the adoption of some theory of mind. Stating that indefinitely many systems of different physical compositions can be conscious, however, runs the risk of being the intuitions of functionalists only—so that smuggling in such a theory of mind which is incompatible with identity theory makes the intuition question-begging. In particular, it constitutes the dialectically objectionable form of begging the question. Polger is saying that there is no better reason for him to accept Standard-MR than there is for him to accept the claim that identity theory is false. Hence calling upon this sort of consideration does nothing to move the dialectic from its starting place about the nature of the mind.

I believe there are a multitude of ways to respond to this charge, which I find to be pretty frankly uncompelling. I will start with what I take to be the obvious move to make against Polger's line, then later consider more indirect responses. This first response is that it is frankly incredible to hold that an acceptance of Standard-MR inevitably presumes functionalism. I am sure there are many computational psychologists, computational neuroscientists, researchers in artificial intelligence, and ethologists who have no idea what functionalism is, but who positively believe that certain physical compositions which satisfy no neurological description, or at least not all the same descriptions, can satisfy the same psychological ones. The grounds motivating their intuition have nothing to do with any preferred theory of mind, but everything to do with the nature of their research and whatever ontological considerations they think that work implies. They are probably functionalists without knowing it, but that does not make the slightest difference. If Polger thinks that their status as computational whatever-ists shows that they are unwitting functionalists, and that accordingly their intuitions can only be question begging against the identity theorist, then he has effectively insulated his view from any possible empirical criticism. I take it—particularly when the view which you have insulated is one that falls within the scope of a science—such a consequence is bad enough to require no real further comment. Obviously it is possible to be motivated by empirical evidence (from one's own work, the work of others, or just a passable understanding of the differences among brains of different species) to hold Standard-MR without a prior attachment to a theory of mind. So even if most of the people who hold Standard-MR are furtively smuggling in their affection for functionalism, that does not spoil the use of the intuition for those who are less partisan. The intuition might be false for all that—and Polger is certainly not wrong to point it out—but that does not make an argument from that intuition to the falsity of identity theory question-begging in the least. Remember, informal fallacies are those defects in arguments that have to do with anything other than false premises.

Polger could perhaps claim at this point that he is not talking about the sort of view that researchers (like those mentioned above) have studied and built up a lot of evidence for—he is simply talking about Putnam's *intuition*, that mere feeling we have that many different kinds of neural substrates can get on with basically the same conscious states (and how this intuition argues against the identity theory). To this extent, I suppose that he is right that simply standing on an intuition like this, without any further support for it than one's affection for a theory of mind, would be question-begging against the identity theorist. But who makes arguments against the identity theorist in that way? Certainly the current literature concerns itself with the scientific case against the identity theory, not the mere intuitive one from Martians, robots, and other fanciful things from the armchair. If it is multiple realization arguments which issue merely from intuitions, motivated only by some subjective dint of plausibility, which Polger finds objectionable, then he can have his point. But since virtually no multiple realization argument (since maybe Putnam) begins from that point, it is not clear why it is worth the bother of attacking them. Modern proponents of Putnam's intuition are not simply infatuated with the view because they love to think they share so much in common with their pets; they think there is major empirical evidence in favor of it which supplies that intuition with dialectical force against other theories of mind. Notice Fodor is not simply belching out a bare intuition in the quote above. His reference to automata surely means he is basing the idea about certain psychological predicates applying to objects satisfying no neurological descriptions on research into artificial intelligence. In this way he thinks the intuition has independent plausibility, motivated by empirical work. So there *is* more reason to accept this premise than there is to accept the denial of the identity theory. It is not just idle speculation about the nature of mental states. This argument is thus not dialectically illicit, or unhelpful. The main premise may be false, but it does not beg the question.

To illustrate this last point a bit more clearly, imagine that I hold the hopeless view that psychological state types are identical to various *geological* state types. Now my opponent tries to meet me with the claim that indefinitely many physical state types can instantiate psychological state types, not (just) geological state types. Does it really cut any ice whatsoever to maintain that such a claim is question-begging against my psychogeologist view? Or, better yet, let us take Polger's view as the opposition to the crazy view; we will call it "Polger's

Intuition” instead of Putnam’s. If his move against the advocate of Standard-MR is worth anything, than the psychogeologist can make the exact same move against Polger’s intuition. He will say, ‘Ah, you have the intuition that only brain state types can be conscious (call it Standard-IT), but the only way in which one can have that intuition is by first finding the identity theory plausible. Of course the identity theory is a rival view to my own psychogeologism, hence you would be taking for granted precisely the issue we would like to settle about the correct view of the mind. That intuition is thus unfortunately question-begging.’ I take it this move would just be outrageous! Certainly no identity theorist would lend it the least bit of credibility. Notice that even the dualist could take the same line—Polger’s intuition, no less than Putnam’s or the psychogeologist’s, is just as dialectically objectionable to the dualist.

Of course what Polger would say is that his Standard-IT is not merely a shot in the dark, fixed only by his predilection for the identity theory. Rather there is a wealth of independent evidence in favor of the claim that it is only neurological kinds which instantiate consciousness. Work in the neurosciences, in particular, offer some such evidence: psychological functions of all kinds are remarkably localized to the same neurological areas in most primates, and sudden loss of any neurological area typically comes at the cost of a sudden loss of the corresponding psychological ability. This shows the ‘intuition’ that only brain state kinds are conscious is not *merely* an intuition. It is a substantially supported claim which happens to imply the falsity of psychogeologism. So the psychogeologist cannot claim to have no more reason to believe in Polger’s intuition than the falsity of his own odd view. ‘Polger’s intuition’ is not question-begging.

But, obviously, everything I have said in defense of Standard-IT goes for Standard-MR. Just as the psychogeologist’s charge carries no weight against Polger’s intuition, the identity theorist's charge can carry no more weight against Putnam's intuition. That intuition, too, is supported by an extensive array of empirical evidence. Again, the burden of the last chapter is to discuss just such evidence, so I will not say anything definite here to give away the game. But so far as it is, Polger cannot simply shrug Standard-MR off as a question-begging starting point. He could try to maintain that so many hold the view only as an intuition, not really supported by any substantial findings in the sciences. But not only will I disabuse him of that error later on, but again it also just seems flatly incredible that Standard-MR's fame in the literature is attributable only to some love affair with functionalism—as if no credence at all could be allowed it without a previous philosophical obligation.

Consider, finally, the response I have been detailing with respect to my insistence that we individuate psychological kinds by adopting a computational picture of psychological processes and demanding that the kinds invoked within the overall process do all of the individuating work. My emphasis has been on appealing to the relevant sciences for establishing sameness of 'mental' state types and differences in the realizer state types. The best going psychological theories maintain that our mental abilities can be described as operations defined over mental representations, so that any creature wishing to share our psychological capacities must also share the same algorithmic processes defined over representations with the same content. Complaining that this is a question begging starting point, however, carries no weight. There is independent reason to believe that these psychological theories are true, regardless of what implications they have for philosophical theories of mind. It is not merely a penchant for computer science that constitutes the sole evidence for the 'intuition' that psychological processes are computational. Psychological models since the cognitive turn have been remarkably fruitful, generating research projects and results at a pace that dwarfs everything before it. If Polger wants to rule out from the very start particular interpretations of empirical evidence, particularly when they offer independent, substantive constraints on theories of minds, because it implies the falsity of a particular view, then he is being completely unreasonable. If one wishes to object to reasoning which advances from empirical premises, one must engage the scientific theory from which the evidence is derived (show why it fares rather poorly as a theory, why its predictions run counter to observation, or why it is poorly integrated with other well established scientific theories, etc.). It is not enough to protest that your opponent's theory implicitly contradicts your philosophical position and is for that reason question-begging. Hence there is no reason for a dialectical complaint against the adoption of these psychological theories.

Recall, however, that I do not really need to offer this kind of response to the possible charge that my adoption of computational theories of psychological processes begs the question. I think it does constitute a serious response to the criticism, but it is unnecessary. I am only using the notion of strong equivalence to provide a rather strong notion of sameness of psychological kind—in order to avoid any granularity issues. I do not require that the computational picture perform any metaphysical work. I will take care of that when I present the empirical evidence for multiple realization. Recall also that I am not defending an inference from 'it is (metaphysically/logically) possible that the same psychological kind is instantiated in different neurological kinds' to 'identity theory is false', rather I am defending the much stronger argument that 'in fact, psychological kinds *are* multiply realized' and so 'identity theory is false'. Reminding the reader of this is another way to see why it should not be a problem that I am adopting a computational picture of the psychological kinds. Perhaps simply adopting such a view allows me to steal the claim that 'it is (metaphysically/logically) possible that the same psychological kinds . . .', but it does not allow me to steal the claim that psychological kinds *are* so instantiated.

But that is beside the present point. All I hope to have done for now is to have provided sufficient reason for ignoring Polger's complaints that these famous arguments beg the question. Fodor's claim about automata, or the 'intuition' people might have about Standard-MR, or my insistence on using psychological theories to guide judgments about sameness of mental state type do not constitute dialectically illicit prior constraints on a theory of mind. So there can be no reason to complain that the well has been poisoned from the very beginning against the identity theory. They are rather independent, empirical constraints. Polger can thus argue that these kinds of premises are false, but it is misleading to claim that the arguments from these premises fallaciously beg the question. They do not.

Section 6.2: Polger on the Nature of Evidence Required for Multiple Realization

The next major theme found in Polger (2009) is an assault on the idea that cases of neural plasticity provide evidence for multiple realization. As I said, I will be discussing exactly this kind of case, so it is worth reviewing Polger's critique. He begins by noting that one of Fodor's and Block's reasons against the identity theory (they call it 'physicalism') is "the Lashleyan Doctrine of neurological equipotentiality" (Block and Fodor, 1972, pg. 160). Polger quickly shoots down the idea that Lashley's doctrine could provide evidence for multiple realization, because—just to cite one consideration—the doctrine of 'mass action' seems to imply that there is only one neurological kind in the cortex which grounds all the psychological types. So we do not get evidence for the one-to-many connection between a psychological kind and neurological kinds. Instead, Polger charitably takes it that Block and Fodor only mean to point out the evidence for how labile the human cortex seems to be. They cite, for instance, the relocation of linguistic function from one hemisphere to the other in cases of brain damage. The big question is thus: "To what extent is evidence of neural plasticity evidence for multiple realization"? (Polger, 2009, pg. 463) Polger claims that a response to this question requires a two-part answer. On the one hand (as I have been discussing in previous chapters on Bechtel, Mundale, and Shapiro), we must come to an understanding of what types of evidence would count in the first place as evidence for multiple realization. Secondly, we must look to see if we find any empirical evidence which fits the constraints we think settle the first question.

Polger's handling of the first question involves borrowing from Shapiro's work, the crux of which we have already considered. I will not repeat the criticism I raised for Shapiro's evaluation of the evidence that would be required to establish multiple realization, but I do want to highlight a few remarks that Polger makes about these constraints that build on the criticism I was advancing. In response to the rhetorical question 'What factors are relevant for deciding whether two realizations are actually distinct?', Polger replies that these answers are "determined by the sciences in question" (ibid., pg. 463). This reply is fairly close to the way I answer the question, but I think there are two problems worth mentioning. First of all, I do not think Polger can endorse the empirical view of determining samenesses and differences of realized and realizer kinds and at the same time endorse what Shapiro says about establishing differences in realizer kinds. Part of my attack on Shapiro was to point out this very problem, arguing that his view would force us to judge certain cases differently than would a view which only appealed to the taxonomies of the relevant sciences (and that this difference in judgment was problematic for Shapiro's view). For instance, Polger tries to explain this appeal to the sciences in the context of corkscrews, with the caveat that we are dealing with the imaginary science of corkscrews. But that is just the problem; there is no science of corkscrews. So we lose one of the main virtues of using scientific taxonomies as a guide: that they provide an independent check on multiple realization that does not run the same risks of arbitrariness as appeals to intuition. What stops me from saying that, in fact, winged corkscrews are actually different in kind from waiter's corkscrews? Somewhat similar to the way jade is realized in jadeite and nephrite, maybe the concept corkscrew is a cover-all term, referring to many different kinds of objects but about which there are no interesting generalizations to make. In other words, maybe there are, as it were, 'species-specific' reductions to make about different corkscrews and that corkscrews, in the general sense, are not multiply realized (in any important sense). Of course if our interest is functional analysis, and we start off with a coarse enough grain (just instantiate the function 'removes cork'), then we can use the differences in our analyses of objects that remove corks in order to show how cork-removers get the job done in many ways. But we are not at all constrained in the case of corkscrews to think that corkscrews are individuated so coarsely—at least not constrained with quite the strength of scientific taxonomies (we do not get to invent those kinds; they are built on their ability to describe adequate laws and predictions). This is at the root of why I suggest we have two different senses of multiple realization: one strict and defined by the sciences, the other imprecise and defined intuitively. So it seems Polger really has to choose one or the other: go with Shapiro or go with the sciences. He cannot use both to figure out what properties are relevant for establishing the realizer and realized kinds.

Secondly, though I do in general agree with Polger's emphasis on the sciences, he seems to neglect the fact that we have to consider both the sciences of the realized kinds *and* the realizer kinds. While advocating an empirical stance, he really only considers the case of corkscrews (about which there is only an 'imaginary' science). He does not seem to notice that, to be in a position to evaluate whether the kind 'corkscrew' is multiply realized, we would also have to say something about the realizer kinds. And this is where the looser, less precise sense of multiple realization really starts to falter. There may be some rough and ready 'science' of corkscrews, but I do not have the slightest idea what science or sciences are supposed to govern the realizer kinds. Sure, there is chemistry and physics, which could go some way toward telling us whether the objects which remove corks are composed of taxonomically different chemical/physical stuff (one is made of steel, the other aluminum). Which are we supposed to consider, however, when determining the empirical extent to which corkscrews are multiply realized? Should we care about the chemical kinds or the physical kinds? The case of psychological and neurological kinds, on the other hand, is much more precise. We have a doctrine which says the former are reducible to the latter, and a response that psychological state types are multiply realized in neurological state types. With corkscrews, we have neither a foil like identity theory nor a clear-cut science articulating the realizer state types. Instead, we are stuck trying to defend some particular intuition about what seems to be the same or different, or, more ambitiously like Shapiro, we must try to formulate some principle which will hopefully provide a non-arbitrary answer to these worries. My suggestion has been to put aside these troublesome cases, but presently I am only concerned with pointing out that Polger does not (or just cannot) make good on his claim about sciences doing all the work—not, at least, when our focus is on corkscrews.

Section 6.3: Polger's Attack on Neural Plasticity

These general sorts of issues about what counts as evidence for multiple realization admittedly run deep and I have addressed some of them already. But now we should move on to Polger's second question: do we have any cases which fit the criteria for being multiply realized? As I said above, a popular place from which to draw that evidence comes from cases of neural plasticity. Polger is aware of it and has a section of his (2009) paper devoted to (re)considering the extent to which plasticity supports multiple realization. I think he is right to devote a full section to plasticity—not many in the literature do more than briefly mention what I take to be a worthwhile neurological topic for philosophers of mind. But I do think some of his remarks about plasticity are misleading (or just incorrect). In particular, I think he underestimates the extent to which the cortex is plastic, and therefore the extent to which psychological kinds are not crucially dependent on particular brain kinds. I also think, owing to how he answers the first question, that very many remarks he makes about plasticity (that is, the interpretation he gives to the relevance of cases of plasticity) are completely wrong. But I want to set those aside for now. I imagine it will be somewhat clear where and how I would disagree. For now I want the focus to be on the empirical status of some of the claims he makes about plasticity.

He begins by dividing up plasticity cases into two main varieties: cortical functional plasticity and synaptic plasticity. The latter kind of plasticity is not typically what philosophers of mind speak about when looking for evidence of multiple realization.[24] This kind of plasticity has to do with the variability of the connections of two individual neurons. The dendrites of one neuron branch off to make connections with up to thousands of other neurons. The connections that are made, however, are not complete. There is a cleft between the connection called the

[24] Though, for a view that cuts against the grain, see Bickle (2003). He thinks a fine-grained analysis of the synaptic level shows that memory consolidation can be reduced to molecular level happenings at individual synapses.

synaptic gap. Instead of touching directly, the pre-synaptic neuron will emit some type of neurotransmitter that acts as a messenger of sorts for the post-synaptic neuron—either by facilitating or inhibiting its firing rate. What is interesting about synapses is that they are the locus of (sometimes) extensive changes. The postsynaptic cell has a number of receptor sites for capturing the released neurotransmitter, the numbers of which can vary greatly depending on many variables. For instance, in the case of drug abusers, the drug of choice will typically act as a facilitator of transmission at particular receptors (in a sense, the drug will impersonate the endogenously created transmitter that the receptor usually receives). As a result of more and more abuse, the post-synaptic neuron will down-regulate the number of receptors for the drug as a homeostatic mechanism. The result is, of course, a greater tolerance for the drug, so that more is required in order to achieve the desired effects. For another example, the famous NMDA receptors only ever work if the right transmitter is present and the post-synaptic membrane is also depolarized (by transmission at a nearby synapse). When the receptor is opened, biochemical reactions take place to up-regulate the number of other kinds of receptors which ultimately strengthen the connection between the two neurons. This mechanism is thought to be responsible for learning of all kinds. Such synaptic level plasticity, while fascinating in its own right, has not typically tempted much philosophical discussion because gross psychological processes are not thought to be correlated with neuron-to-neuron transactions, but rather with whole groups of neurons interconnections and those groups' connections with other large groups of neurons. Also, it is arguable that the result of synaptic plasticity is the instantiation of a *different* kind of function. Advocates of multiple realization, however, want the same function to be instantiated in different neural kinds.

Instead, ‘cortical functional plasticity’ is normally where attention is directed. What Polger has in mind here is fairly broad. On the one hand, there is variability in topographic maps. A curious fact about large areas of the cortex is that neurons that are responsible for processing information about specific areas of the body (or even visual field) are all clumped together in the cortex (and, in the case of the visual field, adjacent receptive fields within the visual field are preserved even in the cortex). Such topographic organization, as it is called, is famously alterable. One cause of its plasticity is the ability of the cortex to reorganize after the loss, e.g., of a particular limb. If a particular finger is lost, the neurons responsible for the finger do not simply stop working altogether. Instead, those neurons are recruited to help process information that nearby neurons process. But the reorganization does not have to be prompted by the loss of the body. Simply doing some particular activity over and over will increase the cortical area devoted to carrying out that activity. For example, very skilled violinists have far more cortical area representing the fingers that touch the strings than a normal person. The topographic map on the cortex is thus not static.

On the other hand, cortical functional plasticity—as Polger intends it—can also refer to the reorganization of the cortex (vis-à-vis its functional specializations) after injury to the brain itself. Depending on the kind and extent of the injury, certain functions are simply lost forever. If one were to have the back of one’s cortex (where visual processing is located) totally ablated all at once, the ability to see is irrevocably lost. But there are many cases where injury (through strokes, tumors, or whatever) do not entail that certain functional capacities are lost—even where the cortical areas damaged were dedicated to carrying out that function. To provide one example which Polger mentions (via Block and Fodor), it is well known that language functions, usually carried out in the left-hemisphere of right-handed subjects, can migrate to homologous areas in the right hemisphere in response to trauma in the left hemisphere. Unsurprisingly, it is this kind of plasticity that usually grabs the attention of philosophers of mind. The idea that certain psychological processes could be carried out by different regions of the cortex naturally lends itself to the idea that psychological kinds are not necessarily in correspondence with brain state kinds. The main point of Polger's discussion, as a defender of the identity theory, is to show that tempting cases like these do not really offer evidence for multiple realization.

His attack on plasticity cases, I think quite rightly, gravitates away from synaptic plasticity. He has a few remarks about why that kind of plasticity is not going to trouble the identity theorist, and I am not going to dispute anything he has to say. In the other case, what he does first of all is to group cases of cortical *map* plasticity (the alterability of topographic maps in the cortex) with other broad cases of functional plasticity (like language lateralization). Then he considers a few different cases of cortical map plasticity and tries to show how they do not meet the criteria for constituting good evidence for multiple realization. What this means is that he overlooks entirely the case for to be made for multiple realization from the cases of cortical functional plasticity that are *not* cases of cortical *map* plasticity.

To mention only one of his examples (to give a taste of his criticism and to set the stage for remarks that he makes that I think are empirically not true), he talks about shifts in topographic maps. In response to these kinds of cases, he explains that "even highly plastic functions do not move about the cortex in a free range manner", but rather even "under normal circumstances . . . the range of plastic change is quite constrained." That is, when we look at the details of topographic shifts in the cortex, we notice that areas representing certain fingers or areas of particular limbs can grow or shrink, but that they never jump around to remote locations in the cortex. So there are cortical changes underlying the 'same' kinds of psychological functions, but it is not clear that those cortical changes are substantial enough to ground the idea that they involve different *kinds* of the cortex. If we want to be exhaustively fine-grained (where a cortical kind is defined to the level of individual neurons), then perhaps there is a case that the same function is mediated by different cortical kinds. But no taxonomy of the brain of which I am aware comes even close to individuating kinds at that fine level of grain, so it is unlikely that any serious scientific kind is going to be classified by being that specific. So if the cortical reorganization is not very substantial—which usually it is not in cases of cortical map plasticity—then it will not count as evidence of multiple realization. So, "it is reasonable to conclude that cortical map plasticity typically involves differences in tokens that do not amount to differences in kind". (ibid., 467-468) In other words, there are differences in the token cortical states that do not suffice to comprise differences in cortical kinds.

I think there are two major replies to make to this line. First of all, what Polger says about small neural changes borders on irrelevance, and, second of all, most importantly, what he says about neural plasticity is empirically false. The problem of irrelevance just involves the emphasis on how small plastic changes are. The size of the changes, it seems, does not bear on the question of multiple realization. So long as the reorganizations involve genuinely different neurological types, they can be as small or massive as you like. Remember, the idea behind neural plasticity is that either new connections are made or different neurons are used to complete the same task. When there is damage to some area of the cortex, the neurons are gone. Besides the hippocampal formation, there is no other area in the entire brain (much less the cortex) known to undergo neurogenesis. So it is not as if the same neurons which performed some task have simply migrated to another location. Even in cases of minute plastic reorganization, new neurons are processing the old information. Polger's point is, of course, that we have to be sure that we have different neurological kinds correlated with the same psychological kind. But strictly speaking this worry is orthogonal to the question of how big or small the plastic reorganizations are. Even the language re-lateralization cases, which strike me as patently not 'small', arguably do *not* involve a genuine difference in neurological kind. As far as I know, the structures in the non-dominant hemisphere which take over linguistic function are always homologous areas—e.g., more or less the mirrored area on the other half of the cortex.[25] To put the point in different words, the quote above where Polger concludes that such cases of plasticity will not amount to differences in kind does not follow only from the premise that such effects are rather small. What Polger, I, and everyone else have to do is show that, whatever differences there are as a result of cortical reorganization, they offer evidence of the same psychological function being instantiated by a different neurological state than was formerly the case. I think cases of plasticity do show this, but I do not think it has much to do with how dramatic or ordinary the cases are. Here Polger, I, and everyone else has got to say something about the correct neurological taxonomy and see how the plastic effects measure up with it. As far as Polger's remarks go, though, he is just using his slipperiness about what the brain state types are to make the smallness point sound much more important than it is.

Anyway, as I said, quite apart from the irrelevance of the smallness issue, I also think Polger is just flatly *wrong* to maintain that cases of plasticity are as restricted as he says. To begin with, the assimilation of what he calls general cortical functional plasticity to map plasticity is not very wise. While map plasticity is usually insubstantial from even a mildly fine-grained perspective of the brain, functional plasticity can be incomparably greater from a mildly

[25] It is worth pointing out that 'homologous' does not necessarily mean 'same'. Indeed, there are more hemispheric asymmetries (vis-à-vis function) than there are symmetries, so it is possible that even if the area in the other hemisphere can process the same information that it should *not* be counted as the same neural kind.

coarse-grained perspective (or even coarser-grained perspectives). Notice that Polger does not talk much about the re-lateralization of language functions (which is curious since this example is Block's and Fodor's main example of plasticity, the attack from which Polger is shielding the identity theory in this paper). Here is an example of plasticity that does not merely involve adjacent areas of the cortex taking over or extending some function; an area completely in the opposite hemisphere is involved. Although, again, we must consult the relevant taxonomies, it is not clear that the same type of defense which Polger supplies for the identity theory is going to work in other cases of functional plasticity. Functions being displaced to bits of the cortex on the order of millimeters away is likely to involve different cortical kinds, much less displacements on the order of inches. The brain is just not that big and neurons are pretty small.

To be fair, it seems Polger is somewhat aware of the crudity of the assimilation. He does consider an example of serious functional plasticity from Shapiro (2004), where he, too, discusses Sharma et al (2000) and the case of ferrets having their visual input redirected at the thalamus to the auditory cortex. I want to talk more about this case, but I would like to do so in such detail that it sprawls out of the scope of this section. So I will spare the reader yet another discussion of the ferrets for now and shelve what I have to say until later. Anyway, this concession does not matter too much because Polger makes sweeping remarks about plastic changes (without qualification) that seem to run directly counter to what you find in neuroscientific journals—to say nothing of the relocation of the language functions which Polger even mentions. He says "even highly plastic functions do not move about the cortex in a "free range" manner" and that, "[u]nder normal circumstances, the range of plastic change is quite constrained." Now, it is not entirely clear what moving about in a 'free range manner' amounts to, what constitute 'normal conditions' is never explained (or even why they matter—does it not count when the multiple realization is artificially created?), and Polger does not comment on exactly how ‘constrained’, or in what ways, he thinks plastic changes are. Worse still, Polger also never gets clear on what the cortical kinds are (aside from hinting earlier that one simply investigates the relevant science). So, to some extent, these remarks might be close to vacuously true. But probably the plausibility of Polger’s defense—taking it that small differences in cortical particulars will not suffice to establish differences in cortical kinds—will require him to hold a fairly strong sense of constraint on plastic changes. As I said above, plastic changes spanning hemispheres or other remote, intra-hemispheric plastic changes are probably ‘unconstrained’ enough to falsify Polger’s claim, and are very likely to scotch his defense of the identity theory.

There are, of course, quite a few examples of the kind of plastic changes that are so unconstrained. First are the examples of plasticity that Polger himself mentions. The case of linguistic functions relocating to the other hemisphere happen, as it were, in the wild, so that type of plasticity does not require very strange circumstances. To some extent, this kind of plasticity is constrained, coming to occupy only homologous cortical structures in the other hemisphere, but it is certainly not constrained in the sense that the function only moves to directly adjacent tissue. Notice also this kind of constraint, if that is what we want to call it, will not give Polger much solace, since a hop across hemispheres is likely to mean not merely irrelevant cortical differences but rather whole differences in cortical kinds.

Next is the case of the cross-wired ferrets. It is obvious that visual information being rerouted to an entirely different cortical lobe (of which there are only 4 in the whole brain) demonstrates the extent to which plastic changes are only constrained in the very loosest sense of the word. The ferrets were ultimately able to achieve a degraded kind of vision from the rewired cortex. The poorer form of vision is interesting when it comes to multiple realization—and Polger rightly jumps on the point—but, degraded or not, the ability to see by using the auditory cortex is a stunning example of the highly plastic nature of the cortex. That an area of the temporal cortex can process early visual information (usually carried out in the occipital cortex), I think, clearly shows that plastic changes can run around fairly freely in the cortex. Or if that is not moving about the cortex in a free range manner, then Polger can only mean that plastic changes have such restrictions in a sense not worthy of interest. He can, I suppose, grant that such changes do not occur in 'normal circumstances'—the re-wirings involved were surgically induced in quite young ferrets. But again it is not obvious why the origin of the lesion, trauma, or whatever that induces plasticity should matter from a philosophical perspective. Assuming we have do have evidence for a psychological process being 'run' by an entirely different cortex, the question of whether that process' being moved there by an act of God or the surgeon's hand is academic. It does not particularly matter how you show the identity theory to be false, so long as you can show that the taxonomies in question are not in a one-to-one correspondence.

At any rate, we are at this point only evaluating whether Polger's empirical claims about plasticity are true. He appears to maintain that plastic changes in 'normal' circumstances are constrained, and it is plausible that the ferret case is abnormal in the relevant way. Not a problem. There are other cases of similar cross-modal plasticity that is not provoked by surgical means. For instance, it is well-known that subjects with early-onset blindness, when reading Braille, have activations within the primary visual cortex. There are classic examples using PET scans (see Sadato et al, 1996) and more up-to-date demonstrations using fMRI scans (see Sadato, 2005). Skipping the details about materials and methods, the basic conclusion is that readers of Braille who were blinded in birth showed activations in primary and secondary visual areas. In particular, while reading Braille, blind subjects showed bilateral cortical activity from the medial occipital lobe to the extrastriate cortex, while during the same tasks normal, control subjects had a decrease in activation. So here is evidence that tactile information is processed, once again, in an entirely different lobe. Like the case of the rewired ferrets, this is another example of the impressive degree of plasticity of the cortex. It also shows that plastic changes can "”jump” to recruit unused but non-adjacent cortical areas" (ibid., pg. 467), which Polger claims never to happen (since the visual cortex never, or stopped, receiving visual information early in life). Also, this example shows that such dramatic plastic changes can occur in 'normal' circumstances, at least where that means the changes are natural and owe nothing to surgical intervention. So surely Polger incorrectly sells short the drastic extent to which plastic changes emerge in the cortex.

Finally, it is worth looking at evidence about cortical map plasticity. It appears that Polger's remarks about plasticity are supposed to hold for cortical functional plasticity more globally, but, even if he is only talking about cortical map plasticity, there is some evidence that contradicts this (more limited) view. There is a case (see Pons et al, 1991) of four cynomolgus monkeys that received deafferentations of the upper limb (3 unilaterally and 1 bilaterally) twelve years before recordings were made in the primary somatosensory cortex at the area which is typically devoted to processing information from the upper limb. This area, about 10 to 14 mm long, is flanked by representations of the trunk and the face. Electrodes were placed in hundreds of locations throughout the area which had been deprived of its typical input. The finding, in all of the monkeys, was a dramatic expansion of the face map, which had 'taken over' all of the former upper limb area. When stimulating (particularly) the jaw are of the monkeys tactually, there was vigorous and organized neuronal response throughout the deafferented region. Thus, while it was generally thought that cortical map plasticity never covered more than a millimeter or two, these monkeys showed that map reorganization can be comparatively massive. Again, Polger never defines what he means by saying plastic changes are constrained, but reorganization of topographic maps on the scale of tens of millimeters provides some decent evidence for thinking that this kind of plasticity could, *pace* Polger, very well involve differences in tokens that do amount to differences in (neurological) kinds. The brain is not very big and contains a staggering number of neurons—so moves across even 1 or 2 millimeters of tissue could very easily cross boundaries defined by neurological taxonomy. Offering no advice on how to type brain states, then falsely pointing out that plastic effects are always small, is not compelling evidence against the idea that neural plasticity offers grounds against the identity theorist.

I hope to have convinced the reader that Polger's attack on the relevance of neural plasticity as evidence for multiple realization is unsuccessful. In particular, I hope to have demonstrated that the extent of cortical reorganization can sometimes be much greater than Polger estimates. Ultimately cases of plasticity have to meet neurological and psychological taxonomy, so I have not settled anything about whether the examples I have mentioned offer evidence of some psychological kind being multiply realized. But I do not think an effective defense of the identity theory is going to come from downplaying the significance of neural plasticity. The brain is capable of reorganizing itself extensively, meaning, I take it, that we have excellent reason for reviewing cases of this kind if we want to settle the question of identity theory or multiple realization. We will have to make sure we have got identical psychological and neurological kinds, of course, but it does seem that neural plasticity has the ability to nix a one-to-one registering of the kinds.

Section 6.4: On Polger's Slipperiness about Neurological Taxonomy

Before moving on to the evidence, I will discuss a final complaint against Polger's attack on multiple realization. When discussing the different forms of multiple realization in his (2002), he takes the question-begging tactic to deal with the stronger forms. When it comes to the weaker, more plausible forms of Putnam's intuition, however, he is forced to adopt a different strategy. Since Weak-MR (at least some creatures that are not exactly like us in their physical compositions can be conscious) and SETI-MR (some creatures that are significantly different from us in their physical composition can be conscious)—at least according to Polger—do not necessitate a previous commitment to functionalism in order to find motivation, he tries to argue that the identity theory can accommodate these forms.

In order to do so, however, he plays on the squishiness of what counts as a brain state kind. This is essentially the same kind of thing Polger was doing in the section above about the size of plastic changes. Since they are so small, the differences in the cortical regions instantiating the psychological function are not enough to constitute differences in neurological kinds. In the previous section I focused mainly on showing that plastic effects are not so small as Polger claims, without grumbling much about the general weakness of this separate line. Here I would like to focus on this other weakness of Polger's attack. Explaining it will, I hope, only encourage an appreciation for the work I have attempted in delineating the neurological kinds, and make the evidence I present in the next chapter more convincing.

The meat of his response to these forms of multiple realization is what he calls the Kim-Adams reply, borrowing from Jaegwon Kim and Frederick Adams. This move is to question whether the differences of the realizing properties (for some single realized kind) are enough for those differences to mark taxonomic distinctions from the perspective of the realizer kinds. He says, "the fact—if it is a fact—that many different kinds of systems can have the same kinds of mental states does not show that they do not all do so in virtue of having some kind of property in common" (Polger, 2002, pg. 150). The same idea holds with psychological kinds as well. It is obvious that, though two events are classified similarly, they need not share all of the same properties; they only need to share those properties that are individuative of the kind. This could potentially be important with respect to Weak-MR and SETI-MR, on the assumption that the differences in the conscious physical compositions are not relevant to the individuation of the physical kinds. From this, Polger is convinced enough to conclude: "But the Kim-Adams Reply shows that identity claims can cover more creatures than is typically supposed—probably all those required by Weak-MR and SETI-MR". (op. cit., pg. 156) So the other stronger forms of multiple realization are question-begging, and the weaker forms are not enough to trouble the identity theorist.

The problem, though, is that Polger does absolutely *nothing* to build the case that Weak- and SETI-MR do not involve taxonomically relevant differences. He admits that probably those who support multiple realization will not be much moved by what he has said and claims that he will build on the case. But the construction of the case only involves an extensive discussion of the 'empathetic' understanding of sameness of psychological kind. Worrying about differences and samenesses at the psychological level, however, does nothing to make the case that Weak- and SETI-MR cases do not involve differences in brain state kinds. In neither his (2002) nor his (2009) does he do anything to make this case. It is true that the move he does nothing to fill in is a genuine worry identity theorists and multiple realization advocates alike must deal with. But simply pointing it out does not constitute an argument—so, as far as I can tell, Polger is not allowed to conclude that the Kim-Adams Reply does anything to establish that Weak- and SETI-MR are not threats to the identity theory.

But is he not aware of all of the strange neurological tales about subjects with extensive lesions maintaining a normal psychology? Is there no reason to suppose that the far-reaching nature of their damage meets the Kim-Adams Reply? Polger makes it clear that he does know about these strange neurological cases. But all he says in response is that:

"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" (ibid., pg. 156)

I take it that this blasé response to strange neurological patients suffers from two obvious flaws. First of all it is a straightforward copout against recalcitrant data. Instead of doing anything like defining brain state types and showing how the ostensibly bizarre evidence is actually accommodated by the identity theorist's view, he just gives up and asserts (not argues) that multiple realization is still not true.

Secondly, more importantly, this sort of copout also implicitly compromises his own view. Of course it helps for some purposes never to be clear on what a brain state type is. If you do not apply any precision, then it is easy to play defense and dismiss any evidence in favor of a view you do not prefer. When in doubt, just remind everyone that we do not know much about the brain. Polger has got a lot of mileage out of just this sort of emptiness about what brain state types are. But if he is right about how empty the notion of a brain state type is, then there is no more reason to accept the identity theory than there is to accept multiple realization. In fact, without filling in a notion of brain state type, *Polger is not even able to formulate what his view is*—for if we really do have no clue how to type the brain states, then there is no telling what the psychological kinds (or conscious mental state types) are identical to. And that means the identity theorist now has no view. So Polger can proselytize all of the skepticism he wants about what strange neurological tales offer by way of philosophical evidence, but he does so at the risk of crippling his own view.

Chapter 7: The Empirical Evidence for Multiple Realization

Finally I want to look at various empirical cases which, I argue, show that psychological kinds are multiply realized in neurological kinds. I have tried to clear the ground of conceptual issues, worrying about how to distinguish the kinds in question or meeting some of the more prominent criticisms of the multiple realiztion argument. Now is the time to see what the empirical evidence says by considering case studies from the neurosciences. Before discussing those cases, however, I want to provide a few remarks about why I mention the evidence I do, remind the reader of an important constraint on that evidence, and say a bit about what work I hope they do for me.

First of all, I am going to opt for more depth to the analysis of cases rather than enumerating as many cases as I can. I think there are many different compelling examples of multiple realization—intra-individual neural plasticity, intra-species neural differences, interspecies neural differences, and even cases of brain-computer interface—but I will only be selecting three different cases and focusing on them. Two of the cases considered deal with interspecies differences and neural plasticity (respectively), so there is some variety in the evidence, but not so much that attention is not paid to the nature of the evidence the cases offer against the identity theory. Another of the examples is the promised cross-modal ferrets case, already famous in the literature because of Shapiro, Churchland, and Polger. I have chosen to focus on it because, as I mentioned above, I do not think the interpretation given to this remarkable case of plasticity is fair. Shapiro and Polger both claim the ferrets do not offer compelling evidence against the identity theory, but I think a more thorough analysis reveals that they do. Since the purpose of this chapter (and the dissertation as a whole) is to discuss the empirical status of multiple realization and the identity theory, reviewing this case again is appropriate.

Where we do look at the cases of plasticity and evolutionary effects, we do need to keep in mind a criticism of Bechtel and Mundale (1999) which I have already mentioned a few times above. They register the objection that usual examples of multiple realization involve an equivocation on the size of grain used to describe the psychological and neurological kinds. Keep in mind that we can find differences in neurological samples at very many levels of particularity. There are gross differences in the locations of the lobes, differences in the organization of layers of cortex, and just minute differences in the numbers of neurons throughout various areas of the cortex. Depending on how fine-grained of a perspective one takes, there will be differences and similarities. At a coarse level, one could ignore various densities of neurons in the temporal lobe, judging two individual tissue samples from this lobe as similar bits of cortex. At a moderately finer grain, one might judge these same sample tissues as different bits of cortex, owing to differences in their distinguishable laminar profiles.

We can make the same point with respect to psychological processes. We could start with a very coarse-grained perspective and treat as alike any two creatures who can see (smell, hear, etc.). At a more fine-grained level, we might distinguish these same creatures because of some greater visual acuity in one. Most birds, for instance, have significantly better vision than humans. If our concern when individuating vision is to distinguish between creatures which can see flying insects from many yards away, then we will probably find reason to differentiate between human and avian vision.

The reason behind pointing out this variability of grain size is to avoid any trivializing mismatches of grain. One could, for instance, take a very coarse-grained view on the psychological side (classifying as the 'same' anything which can see), while matching that judgment with a very fine-grained view of neurological taxonomy (only grouping together bits of tissue, say, that contain exactly the same number and kind of cells). In a sense, one would have shown that (a very broad notion of) vision is multiply realized, but this kind of multiple realization is trivial and cheap. Identity theorists will not be too much troubled if that is the best evidence for multiple realization; at such a mismatch of grain size, there will be samenesses through differences every time. The mismatch problem works the other way, too. Rather than involving a far too coarse notion of the realized kind, one can be far too particular at the realizer level. The human cortex, for instance, is famously different across all individuals, yet there do seem to be interesting generalities to state about the cortex despite these differences (we all have four lobes, e.g.). Or, to use the more vague examples, if we build two carburetors to function exactly the same way, made of exactly the same material, there are bound to be minute differences in the atomic compositions of the two devices. We would not intuitively be inclined to say those carburetors are multiply realized, for something like opposite reasons as the case of human and avian vision. The level of grain at which the carburetors are judged the same might be reasonable, but it is starkly out of sync with the incredibly fine level of grain one would have to use in order to detect any differences between the individuals.

Again, this example of multiple realization seems to be trivial and insignificant. What the identity theorist holds is more accurately like: psychological kinds are matched up one-to-one with neurological kinds, where the grain level at which we individuate the kinds is not wildly mismatched. The advocate of multiple realization must surely be sympathetic to this line, too: the real guts of the idea is that the exact *same* kinds of things are realized in significantly *different* kinds of things. Anything short of that feels ultimately like sleight of hand. So there is good reason for fussing about how coarse or fine-grained one's perspective is when one classifies the kinds to be compared, and taking care that the perspective at one level is at least roughly preserved at the other level.

One of Bechtel's and Mundale's objections is that many examples of multiple realization fail to be interesting examples because of a mismatch in granularity. Putnam's famous illustration against the identity theory about octopus and human pain perhaps provides an illustration. Since we know nothing about how exactly pain 'feels' in octopi, it is possible that, when comparing with human pain, we are simply adopting a loose, coarse-grained perspective for classifying pain, then matching it with a more fine-grained perspective on the neurological structures underlying each creature's pain. Identity theorists—not implausibly—love to point out that, if we could get into the skins of octopi, we would probably judge the pain to feel different. Though this line perhaps runs the risk of the opposite mismatch of grain (fine psychological typing with a comparatively more coarse neurological typing), the point probably still stands with respect to Putnam's example. It either begins with dissimilar kinds of pain (so the example never gets off the ground) or adopts a grain size that is not maintained when judging the distinctness of the neurological structures. Notice also that there is nothing objectionable about maintaining a fairly coarse level of grain at both ends of the taxonomy. But, vis-à-vis Putnam's example, keeping that coarse level of grain might also mean that we should judge the neurological structures in humans and octopi to be the same, or at least not relevantly different—once again resulting in a situation that provides no evidence for multiple realization.

For my own part, I think Bechtel and Mundale have a point worth keeping in mind. One might think that using the sciences to fix the classifications would lock the level of grain, so that all we really had to do was look at what the psychologists and neuroscientists were talking about and we could evaluate multiple realization from there. The sciences, it might seem, are blind to grain size issues; they simply classify things the way they have to in order to make all the generalizations they can. But it is important to note that the sciences can also be more or less fine grained with their own perspectives. There are variations within each science about the level of specificity used. For example, one could talk broadly about declarative memory, or more specifically about episodic or semantic memory more specifically (two different types of declarative memory). Similarly, one could talk about vision in general (across insects and primates), without bothering to drag through the details that differentiate between the visual capacities across species. So there is some reason for keeping the grain issue in mind. At least in the interest of fairness I will make sure to discuss the level of grain used in order to allay any fears that my cases are only trivial examples of multiple realization. I will try to show that some congruence exists between the levels of grain used to distinguish the kinds in question. But there will also be a case where grain sizes make a significant contribution to the discussion. Particularly with the case of the rewired ferrets, I want to use this appeal to grain size (among other strategies) to revive this case as evidence in favor of multiple realization—despite its fame in the literature as a failure for providing that evidence (in Shapiro, 2004 and Polger, 2009).

Lastly, before getting on with cases, I want to mention that none of the cases I mention really provide knockdown evidence against the identity theory. Since the question is ultimately an empirical one, no evidence could be. We are unfortunately at the mercy of scientific theorizing. This helps to guard against slippery moves that, for instance, Polger likes to make in twisting the evidence presented in empirical cases toward the identity theory, but the scientific sword's independence from philosophy also makes it double-edged. If 'ideal', completed neuroscience (or psychology) looks far different than what we have now, the evidence here is probably undermined, too. As everyone warns, if the completed sciences bare little resemblance to what we have now, all bets about identity theory/multiple realization are off. I present what I take to be the best evidence we currently have against the identity theory, evidence which I think is on the right track, though obviously researchers could still make important discoveries which change the philosophical landscape completely.

There are also well-known problems with some of the evidence I use. Just to mention one particular kind of evidence, PET and fMRI scans are somewhat notorious for the indeterminacy of the information they convey. Roughly speaking, they both look at metabolic activity in the brain, supposing that the more energy an area of the brain consumes, the more important that area is for processing the information necessary for completing whatever task was assigned during the scanning. But the statistics used only provide average differences among (sometimes large, therefore somewhat indiscriminate) chunks of the cortex, where significant differences can be as low as two or three percent increases in metabolic activity. The results, since each individual's brain is different, are also normalized and mapped onto an ideal brain, thereby abstracting a bit from any idiosyncrasies. Finally, these scans really only provide us with correlations. Processing certain types of information, as far as we can tell, is correlated (more or less) with an increase in metabolic activity in certain areas of the cortex. These results do not tell us that those areas are definitively the substrates of the processing; it is still left open, say, whether the areas identified are only necessary precursors to the actually important areas.

In response to these kinds of worries about the evidence I use, we should keep in mind that fMRI scans and the like constitute the best available tools for gathering data that we currently possess. Though that evidence is certainly defeasible, we should not simply turn up our noses with Fodor and disregard what, for instance, the scans have to tell us. A healthy dose of skepticism is usually a good thing, and we should be willing to restrain our enthusiasm about what fMRI scans demonstrate. But insisting on too much skepticism in the face of the best available evidence is just flatly unreasonable. If, using the best existing means, we feel like the resulting data supports a position, we have strong reason for accepting that position (even while minding all the relevant caveats). Anyway, I will try to be as frank about how strong the evidence is (to the best of my knowledge about it), and will warn the reader where the weaknesses in my discussions occur. I think, in the end, that the straws blowing in the wind favor the claim that psychological states are multiply realized, though it is worth confessing up front that nothing like a demonstration has yet occurred.

Recall, in closing this introduction, that the main idea is to show that the same psychological kinds are instantiated in different neurological kinds, where the neurological kinds in particular are typed by Brodmann areas. The cases I discuss below hopefully provide evidence that the same kinds of psychological state types are instantiated in different Brodmann areas. If I can do that, then there would be evidence for thinking that psychological kinds are multiply realized in neurological kinds, which in turn would show that the identity theory is false.

Section 7.1: An Example of Cross-Species Multiple Realization

In honor of Putnam's classic line, I want to look at least one case of a single psychological kind which is multiply realized across two separate species in distinct Brodmann areas. This is somewhat problematic in the case of monkeys and humans, however, because there are known differences in the maps—particularly in the frontal areas where natural selection appears to have provided advancements in humans. Nevertheless I look at the same psychological kind and how it is realized in different Brodmann areas in monkeys and humans, though that is not going to cause a problem because, remember, I am not trying to say the same psychological kind is realized in the same Brodmann area. So long as there are important cytoarchitectural differences between the areas in monkeys and humans which realize the psychological kind in question, then the kind must be multiply realized.

7.1.1. The case study: what the researchers were looking for and what they found

The psychological kind in question is *spatial working memory*. The principle paper I want to look at, though I will cite a few more, is Courtney et al (1998). They investigate an area in the human frontal cortex that is specialized for spatial working memory, though this area is different from the area in monkeys which specializes in same sort of processing. Hence it offers evidence in favor of the view that the same psychological kind is multiply realized in different neurological structures across species. I will take a look first at what it is Courtney and colleagues were looking for, then I will look at the experimental design used to collect the relevant data, and finally we will see how they interpreted the data. We can then examine the implications of these findings for multiple realization.

In both monkeys and humans, the prefrontal cortex has been implicated in processing working memory in general. Numerous experiments have shown that, during working memory delays, sustained activity takes place in the prefrontal cortex (see, for example, Goldman-Rakic, 1995; Cohen et al, 1997; Courtney et al, 1997). In monkeys, there is a clear delineation between spatial working memory and working memory for patterns, colors, and faces, with the former kind of working memory having been localized to the principal sulcus. On the other hand, the results of brain imaging studies in humans, up to the time of the paper we are considering, had been more ambiguous. Since the area for spatial working memory lies in Brodmann area 46 in monkeys (Courtney, et al, 1998, pg. 1347), most trials had focused on that Brodmann area in humans. While area 46 is activated during exercises for spatial working memory, performing other working memory tasks (e.g., verbal and visual tasks) also activates the area. So the existence of a specialized area for strictly spatial working memory in humans had been questioned.

The point of Courtney et al was to provide evidence that there *is* a specialized area for spatial working memory in humans, "but it is not in BA 46" (Ibid., pg. 1348). Instead, they supposed that this area, if it did exist, would be slightly anterior to an area known as the 'frontal eye field' (FEF), because in monkeys the area for spatial working memory is located just anterior to the FEF. In humans, the FEF has migrated to a more superior and posterior location, meaning that regions more anterior to it would be found in the vicinity of the superior frontal sulcus, a rather large indentation in the higher regions of the frontal cortex in humans. Courtney and colleagues report that this area has demonstrated activity during working memory delays in other studies, though its proximity to the FEF and premotor cortex seemed to indicate a non-mnemonic role for activity in the superior frontal sulcus.[26]

In order to determine whether the superior frontal sulcus is specialized in spatial working memory in humans, they devised a spatial working memory task and used fMRI to look at how

[26] In what follows I rather thoroughly explain the experimental design, the results, and how the results support their conclusion. If the reader wishes to take the methodology for granted, not much would be missed by skipping the details and moving straight to the evaluation of this study as evidence for multiple realization.

much energy the area was consuming during the task. They tested for four separate criteria, which together define whether the area is 'specialized' for the task:

"(i) the area must show sustained activity during spatial working memory delays; (ii) such sustained activity must be greater during spatial working memory delays than during delays in other types of working memory tasks, in this case working memory for faces; (iii) the sustained activity during spatial working memory delays cannot be attributable to preparation for a motor response, which would indicate a premotor rather than a mnemonic function; and (iv) the area must be distinct from the FEF" (Ibid., pg. 1348)

If the activity in the superior frontal sulcus met each of these four criteria, then that would constitute good grounds for saying that activity there was devoted to spatial working memory. This is exactly what they found.

In order to show the first criterion was met, they had 11 healthy human volunteers perform a particular task. The experimental design was actually quite complex, mixing all the control elements into one particular task. To begin with, there was a sensorimotor control task used to make sure the frontal activations for the working memory tasks were genuinely involved in memory processing. This task used three scrambled pictures of faces, each presented on a screen for two seconds, followed by a nine second delay before finally subjects were asked to provide a control response of pushing two buttons. Then, after another six second intertrial interval, a new set of stimuli are shown on the screen. This time, the experiment actually addressed working memory. The screen would show three faces in three different locations, each face-location pairing appearing for two seconds. After a nine second memory delay, a test stimulus of a particular face at a particular location appeared on the screen. Some subjects were asked to remember the locations of the faces on the screen, then asked to press a left or right button indicating whether the test stimulus' *location* matched any of the locations of the three stimuli in the memory set. Alternatively, given the three faces in different locations, subjects were asked to indicate by a left or right button press whether the test stimulus' *face* matched the faces of any of the three faces presented in the memory set. Obviously, the idea behind the second set of stimuli was to test for neural activity during working memory. The task called for monitoring locations and faces because the experimenters wanted to make sure of the dedication of activity within the superior frontal sulcus to spatial working memory, and not just working memory in general.

Since the first criterion was simply to show sustained activity in the superior frontal sulcus during specifically spatial working memory delays, they had to find a way to measure neurological activity in this area during the tests. To do so, they used fMRI scans and looked at voxels that were significantly activated by each component of the task. In the first case, the entire dorsal (higher, upper part of the) frontal cortex (the area where the superior frontal sulcus, among other things, is located) showed sustained activity during working memory delays in each of the eleven subjects. However, within the dorsal frontal cortex, 66% of that sustained activity across subjects is found in the superior frontal sulcus. Just for a contrast, 69% of the transient, unremarkable activity in the dorsal frontal cortex occurs in the precentral sulcus, which is just lateral and posterior to the superior frontal sulcus. The P-value for the difference between the localizations of sustained and only transient activity within the dorsal frontal cortex was less than 0.002, so it is rather unlikely that the percentages were anomalous. The contrast seems to be enough to conclude that the first criterion is established: the superior frontal sulcus does demonstrate sustained activity during working memory delays.

The second thing to show in order to demonstrate the area's specialization is that the superior frontal sulcus did not show sustained activity during other types of working memory delays. This, of course, was the point of adding in the task for identifying faces. Of seven of the participants, the superior frontal sulcus showed significantly more sustained activity ($P < 0.01$) during spatial working memory delays than during face working memory delays. Again, for a contrast, the left inferior frontal cortex showed significantly more sustained activity during face memory delays compared to spatial working memory delays. Areas in the middle frontal cortex, more posterior and anterior than the superior frontal sulcus, also showed significantly more activity for facial working memory than for spatial working memory, while even lower in the ventral frontal cortex activity did not reach significance for a difference in the two working memory tasks. So it appears, with respect to the inferior frontal cortex and the superior frontal sulcus, that the results show a double dissociation of function. While the former shows higher haemodynamic response for face working memory tasks, the latter shows a higher response for spatial working memory. So the superior frontal sulcus passes the second criterion, too: the sustained activity there is greater for spatial working memory than for other working memory tasks (at least face working memory, in this case).

Also, these results show that criterion three is satisfied. Both the face working memory task and the spatial working memory task involve the same button-pressing responses after the test stimulus. If the activity in the superior frontal sulcus were attributable to (preparation for) a motor response, then one would expect virtually the same level of activity there in both kinds of working memory tasks—i.e., even in the face working memory task, the subject must push the same buttons as he or she does in the spatial working memory task. That there exists a significant difference in the levels of activation for either tasks shows that the response in the superior frontal sulcus has nothing to do with the motor areas of the brain. So the third criterion is also met.

To establish the fourth criterion, that the area involved in spatial working memory must be distinct from the FEF, Courtney and colleagues had four of the subjects perform a final task, involving visually directed eye movements. There were fifteen second blocks of guided saccades on a screen, mixed in with fifteen second blocks of a blank screen, where participants were asked to look at the center of the screen and try not to move their eyes. As had been established elsewhere (Petit et al, 1997; Luna et al, 1998), the area of the greatest activity in the whole dorsal frontal cortex was found in the precentral sulcus. In fact, a full 85% of the activity in the upper half of the frontal cortex in response to the eye movement task was found in that sulcus—distinguishing it well enough from activations in the superior frontal sulcus. Furthermore, the face identification task also involved the same kinds of eye movements as the location identification task. So the significantly lessened neural activity in the superior frontal sulcus in response to the face working memory tasks shows that activity there has nothing to do with oculomotor control. So the fourth criterion seems to be established for the response in the superior frontal sulcus as well.

Ultimately, the neurological areas for spatial working memory and the frontal eye fields are located slightly more posterior and superior in the human brain (than in the macaque brain), though they do bear the same relationship to one another as in monkeys, with the area for spatial working memory just anterior to the frontal eye fields. Probably the reason for the displacement of these functions in the cortex has to do with the evolution of other areas in the human cortex—either by the development of brand new functions, or by older functions being greatly exaggerated in humans.

7.1.2. Evaluating the Case Study's Evidence for Multiple Realization

What is the import of these results for multiple realization? As ever with the empirical literature, the philosophical side is not exactly stressed. Researchers have their goals, and philosophers have their own. In this case, there are two issues to be concerned about: whether the psychological kinds involved are actually the same, and whether the realizers for the psychological kinds are genuinely distinct from a neurological perspective (viz., from Brodmann's perspective). If we get a pair of affirmative answers for this case about spatial working memory, then we have evidence in favor of multiple realization. We should also discuss Bechtel's and Mundale's grain issue. If it turns out that the interpretation of the data depends on a mismatch of grain size when individuating the kinds at each level, then the philosophical importance of the case is undermined. So we need to make sure there is no slippage between the stringency with which we are approaching either level.

The hard case to make is for the psychological similarity of spatial working memory in humans and monkeys. The present study does not draw any conclusions (nor does any other study I know of) about how spatial working memory is similar across all primates—or, more to the point, how this ability is similar in macaques and humans. For one thing, the *tasks* involved are more or less always the same. When examining working memory (of any kind you like), the typical design involves some sort of delayed matching task, where a sample object is presented, then perhaps masked with other distractor images presented, finally the sample is presented again, and the subject must push a button, or look in a particular direction, indicating that the sample has reappeared. Courtney and colleagues use a design like this, and other papers they cite have similar designs (see, e.g., Wilson et al, 1993)—though the former use human subjects, while the latter uses only macaques. So at least in terms of design no differences are expected about what monkeys can remember and what humans can.

The real question, though, is whether either memory system has the same, so to say, acuity. Importantly, we do already know that monkeys and humans use roughly the same kind of algorithms to process the information. Though the neural correlates of the abilities have grown apart, humans and monkeys alike divide memory processes in roughly the same way, separating working memory from long-term memory, and even facial working memory from spatial working memory. The basic flow of information is the same across both primates. But what would be nice to know is that humans and monkeys are both about as good as each other in terms of spatial working memory—that both can remember events with the same sort of spatial content, for about as long as each other, about as much content as each other, etc. If monkeys, for instance, qualitatively have much worse spatial working memories, then we will at best have to adopt a more coarse grain of kind individuation (in order to judge the capacities to be the same across both monkeys and humans), a move which will force us to adopt a large enough grain size neurologically that we jeopardize the possibility of the neurological kinds being genuinely distinct (from that rough grain size). That lack of similar acuity would offer some reason for thinking that the algorithms were not perfectly similar.

Unfortunately, I am unaware of any such study, and would not be too surprised if no one had bothered to investigate the question. I can point to another working memory study, however, which does show monkeys perform rather well at least for that study's particular task. In Freedman et al (2001), monkeys had to perform a delayed matching-to-category task, where a sample was presented of a particular category, and a test stimulus was presented requiring the monkeys to respond whether the test stimulus matched the sample in terms of belonging to the same category. The categories involved were clear-cut in some cases, but very blurry in others—the whole idea of the study was to explore the neural basis for 'categorization', the ability humans and primates possess to group certain stimuli into the same group, though there are physical differences between group members (e.g., apples and bananas are in the same group, though they differ in many properties). In cases where the samples were clear cut, performance at matching was near 100%. At the boundaries, where the samples were mixed and not obviously belonging to a category, performance was still at 90%. This is, of course, not a spatial working memory task, but it does illustrate that the monkeys in the study were able to remember the stimuli with high accuracy. Even if humans could perform better (working at, say, 95% accuracy in the mixed cases), they could not do so much better that we would want to say human working memory was simply of a different *kind* compared to monkeys, or that in order to judge them as similar we would have to take a liberally coarse level of grain. Keep in mind that in none of these studies do humans or monkeys have much time to make a decision. The sample stimulus will be up for around a second or so, followed by varying delay intervals, and the test stimulus will also only appear for around a second or two. Neither monkeys nor humans have much leisure in deciding on a match. They do not particularly need such leisure, but it should be pointed out that the tasks typically do put stress on the subject's memory.

I know these remarks do not really clinch the case that spatial working memory in humans and monkeys is identical. But it is telling, I think, that experiments with either species rely on virtually the same design every time, and there are certainly studies where monkeys perform remarkably well in terms of accuracy. These are definitely the sorts of facts we would expect if we were dealing with the same kind of spatial working memory. Assuming that there is a fairly sharp comparison to make there, we would, for instance, not want to change the experimental design. If a particular design seems to churn out good data (it takes care of all the relevant controls, there are no outstanding confounding elements, etc.), experimenters will want to stick with it. It is a small vote of confidence for the similarity of macaques and human working memory, then, that the same experimental designs are used for both.

To repeat, though, it is assumed that the processes humans and monkeys use for remembering things, regardless of the acuity of their spatial working memories, are rather alike. The same sorts of dissociations one finds in amnesiacs (say, between working memory and priming effects) are also found in monkeys. In this case, the dissociation between spatial and facial working memory is predictable from the dissociation in the visual cortex of humans of the dorsal and ventral streams (which keep track, respectively, of information about *where* an object is and *what* an object is). That is, the dorsal and ventral streams are preserved even as far as the frontal cortex, where working memory also has a dorsal/ventral split between spatial locations (where) and faces (what). So, to the extent that there is a spatial working memory in humans and monkeys, there are grounds for thinking that the memory processes in either species use at least roughly the same algorithm. And that counts as one more reason for thinking the two spatial working memories are alike in acuity.

What about the second worry: that these two neurological substrates are genuinely distinct? In terms of Brodmann areas, the neurological taxonomy I have argued we should adopt, we *are* dealing with neurologically distinct substrates. In macaques, spatial working memory is found in Brodmann area 46, while Courtney and colleagues claim that the area in humans is "relatively more superior and posterior than a similar functional region in monkey frontal cortex [but] has a different BA [Brodmann area] designation" (Courtney et al, 1998, pg. 1348). So they do not exactly say in what Brodmann area spatial working memory is supposed to be found, but it is easy to look at what Brodmann areas the superior frontal sulcus occupies to find out. It seems to take up Brodmann areas 8 and 9 in humans. Brodmann area 8, in particular, is more posterior and superior than area 46, and it is the location of the FEF in humans, too. So that is a good bet about which neurological kind is involved in this case. Whatever the case about Brodmann, it is at least clear from what the researchers say that the neuroanatomical areas involved are distinct for monkeys and humans. In humans, as mentioned, the superior frontal sulcus is the all important area; but in monkeys, the principal sulcus is the location of spatial working memory. At any rate, even if Courtney and colleagues do not identify which area specializes for spatial working memory in humans, they do say it is in a *different* Brodmann area than macaques—and that makes sense of why no other researchers were able to locate such a specialized area in humans. Hence it seems like the neurological correlates of spatial working memory in humans and monkeys are genuinely distinct neural kinds.

What about the grain issues? From a psychological kind standpoint, I think the level of grain is fairly fine. To begin with, we are not concerned with a rather coarsely typed psychological category. The research here does not have to do with memory considered as a whole. It does not even have to do with *working* memory as a whole. Rather, the research is limited to revealing something about *spatial* working memory. There is a dramatic difference between these three things. Memory as a whole is the ability to store, retain, and recall any kind of information. Computers have a memory in that sense. Working memory, to be a bit more specific, refers to the ability to store immediate information related to goals the organism is trying to achieve, even in the face of continuous distraction. The goal of working memory is not to store long-term information, but rather to track certain information just long enough to make goal-oriented action possible. In this sense, the responsible areas of the brain work like a scratchpad, writing in information that is important now, then erasing it and rewriting new information that becomes salient later on. One can, for instance, lose the ability to store information in the long term, while retaining short-term, working memory. Most serious amnesiacs have lost their ability to form long-term memories, though their working memory is fully intact. So they remember semantic, visual, and spatial information just so long as they are not distracted by other goals or information. Hence these sorts of subjects will meet the same people repeatedly over their lives—or even just throughout a day—and never remember having met them. The moment those people leave the room, and working memory becomes devoted to storing other information, knowledge of having met them is gone forever. More specific than that, however, we are concerned with the *spatial* variety of working memory—that is, the ability to store information in the short-term about where some object or objects are located. We can distinguish this kind of working memory, as we have already seen, from working memory about faces, or about words. So the level of grain applied to the psychological kind is rather fine. There are, of course, worries about how similar spatial working memory is across species, so some of the exactness of grain is arguably lost. But, on the whole, the grain size is rather fine.

And what of the grain size for individuating the neurological kinds? Here, using Brodmann areas, I think the grain is fairly coarse. Brodmann's areas are identified via cellular differences, differences in cell densities, and other specific features, but the results only provide 50 areas in a brain with about 100 billion neurons. The differences between the different areas, too, are usually quite stark. It is not as if a slight difference of cellular density at one level is enough to warrant a different Brodmann area. Typically, the different areas involve the presence of completely different neurons and/or massive density differences at each of the six layers. On the basis of the cytoarchitectural features Brodmann employs, we could be much more selective in carving up distinct subsets within a single Brodmann area (in fact, there are such subdivisions within area 23 in macaques). So the neurological grain size, while perhaps appearing quite fine at first, is arguably somewhere in the middle of coarse and fine.

Ultimately, I do not think there are good grounds for complaining about a mismatch of grain. Even if I am right in saying that the psychological grain is cut on the fine side (while the neurological grain is neither fine nor coarse), there should be no objection. Being super-specific about the psychological kinds, then matching those kinds with a comparatively coarse neurological kind, seems to provide the best evidence for multiple realization. It should be that coarse individuation of neurological kinds means many different psychological functions will simply get lumped together neurologically. If we are super fine psychologically, accepting only very, very alike individuals as kind-identical, matching those kinds with rough-and-ready neurological types, and we *still* have the same psychological kind matched with really different neurological kinds, then the identity theory looks in serious trouble. At any rate, the really objectionable match-up, as Bechtel and Mundale have shown, involves the opposite mismatch: coarse psychological grain and very precise neurological grain. Memory, conceived really broadly, is probably multiply realized in every single human (not even to mention other species, or computers, too), assuming you match that kind up with a demanding neurological taxonomy. No two hippocampi, I take it, are exactly alike. That sort of multiple realization is rightly called trivial. But I do not think this case about spatial working memory makes that mistake. At worst, both levels of grain are the same. More likely, the psychological grain is a bit finer. But, as I have pointed out, that *boosts* the case for multiple realization; it does not undermine it.

7.1.3. Conclusion

So I take this case about spatial working memory to constitute fairly good grounds for saying the same psychological kind is multiply realized in distinct neurological kinds. Humans and monkeys both process memory in the same way—dividing up working memory from long-term memory, spatial working memory from other kinds of working memory—and do so with probably roughly the same acuity. In addition, they also realize these abilities in distinct Brodmann areas. Since I have argued that the neurological taxonomy to adopt is sensitive to Brodmann's areas, we have reason to say spatial working memory is multiply realized in humans and monkeys.

Section 7.2: An Example of Multiple Realization from Neural Plasticity

Next I want to look at neural plasticity. The case I focus on is a smaller instance of a more general phenomenon which, as far as I know, is new to the literature on multiple realizability. This phenomenon I will boringly call the Gradual Reorganization Principle (hereafter GRP). It states that the brain will reorganize neurological processes in response to degenerations of all kinds, thus saving psychological capacities, assuming sufficient time is allowed for the brain to adapt. As the brain damage slowly takes away neural functioning, the brain will recruit other regions, sometimes remote and sometimes near, to perform the processing required to maintain the psychological capacity. The length of time required by the brain to redistribute the processing adequately is difficult to determine, but there is a clear difference in functional restoration between abrupt brain damage (e.g., from strokes) and gradual brain damage (e.g, from slow-growing tumors), where the former kind usually results in marked psychological deficits and the former in little or no measurable deficits. This principle is relevant to multiple realization because neurological cases which can be described by it provide evidence of the same psychological abilities instantiated in a variety of neurological substrates.

But rather than speaking generally about the GRP, it would be more helpful to look at a specific example of the principle: the so-called *serial lesion effect*. Researchers have known about the serial lesion effect, conducting numerous experiments in order to pin down the particulars, since at least the 1930s (see Kennard, 1938 or Kleitman and Camille, 1932) and the 1940s (see Ades and Raab, 1946, and Kennard, 1942). In fact, Finger et al (1973) report it was known that resection of neural tissue in stages, compared with resection of the same tissue in one surgical sitting, is correlated with lower mortality rates from as far back as the middle of the nineteenth century. Fluorens (1824) was able to keep decorticated pigeons alive for much longer periods of time when the tissue removal was done in small amounts rather than immediately. Surprisingly, however, this effect has not made it into the philosophy of mind literature, though it offers excellent grounds for doubting the identity theory. The basic concept is easy to understand. To phrase it informally, the idea is that removing neural tissue in a serial manner, bit by bit rather than instantaneously, usually allows for a substantial degree of functional recovery—just as the GRP predicts. In other words, if one compares two brain damaged individuals, both of whom have lost the same amount of tissue in the same locations, the individual whose damage was received in a piecemeal fashion, rather than being immediate and all at once, will enjoy a much greater degree of psychological recovery than his all at once companion. Many times the patients with seriatim damage still do have noticeable, though usually slight, deficits in comparison with non-damaged controls. But in many other cases—the really interesting ones from the point of view of multiple realization—serial patients have no measurable psychological deficits at all, compared with neurological normals or whomever else. These later cases, where brain damage is paired with no functional anomalies, provide just the sort of evidence an advocate of multiple realization would be looking for. Here are the same psychological abilities linked with fundamentally different neurological structures. Remarkably, where certain neurological kinds were previously critical for instantiating psychological kinds, after serial lesions to the previously important tissue, the psychological kinds are relocated to different tissue, hence to different neurological kinds.

7.2.1 A Case Study of the Serial Lesion Effect

An illustrative study of this effect is Patrissi and Stein (1975). In this experiment, the researchers had forty-six albino rats placed into five surgical treatment groups. One of the groups underwent bilateral removal of the frontal cortex, removing both hemispheres of the frontal lobe in one surgery. Three groups received only unilateral resections of the frontal lobe in two separate surgeries, one hemisphere being removed during the first surgery and the other removed in the second. In these groups, the one half was removed followed by the other half either 10 days later, 20 days later, or 30 days later. The final group served as the control group, undergoing only sham surgery. From there the different groups were tested in a T-maze on a spatial alternation task, where, after the rats are rewarded for choosing one of the goal arms, the reward was switched to the other arm on the next run. The goal was to ‘reach criterion’, which meant correctly choosing the reward arm as it alternated for two straight days (at 16 trials a day).

The results were measured in how many days it took each group to reach criterion on average. In terms of raw data, the sham surgery rats took 124.8 ± 35.22 days; the two-stage, 30-

day interoperative interval rats took 140.8 ± 32.7 days; the two-stage, 20-day interoperative interval rats took 156.4 ± 48.58 days; the two-stage, 10-day interoperative interval rats, however, took 270.4 ± 56.17 days; finally, the one-stage rats took a whopping 427.4 ± 67.7 days to reach criterion. In terms of analysis, the shams, the two-stage, 30-day interval rats, and the two-stage, 20-day interval rats did not differ significantly from each other. In other words, the differences between the sham rats and the one-stage lesion rats were—unsurprisingly—significant, the latter group taking over three times as many days to reach criterion as the former group. The same results were obtained comparing the sham group to the two-stage rats which were only allowed a 10-day interval between surgeries. The difference there was only about 130 days more for the seriatim group, but it was enough to reach significance. In fact, these two groups (one-stage and two-stage, 10-day interval) were significantly different from the other three groups. Now, Patrissi and Stein do not provide the exact p-values for each of these comparisons (claiming only that the mark required for significance was set at the customary 0.05 level), but it is worth noting that the comparison between the shams and 10-day interval rats must have been closer to *in*significance than the sham/one-stage comparison, suggesting that even the two-stage, 10-day interval seriatim rats were much more psychological similar to the sham rats than were the one-stage rats. Indeed, Patrissi and Stein report that the one-stage rats did perform significantly worse than the two-stage, 10-day interval rats. This fact further underscores the importance of the GRP: the brain will recover from lesions if the damage is gradual, even in cases such as the two-stage, 10-day rats where the interoperative interval is relatively small.

Most importantly, however, the sham group did *not* differ significantly from either the 20-day or 30-day two-stage lesion rats. This indicates that the differences in the mean number of days to reach criterion for these three groups were likely to be statistical anomalies and not supportive of any serious psychological difference. Consequently, there are no measurable psychological differences between the two seriatim groups with extensive neurological damage and the sham group with no neurological damage at all.

In addition, once the previous testing was completed, the authors decided to add additional experimental animals to control for the possibility that the surgical trauma might have accounted for much of the delays in reaching criterion. So one group of seven rats received a two-stage sham operation with a 10-day interoperative interval, and a second group of five rats simply received two separate injections of saline 10 days apart to control for the injections involved in all the other operations and for trauma simply during handling. The results for these rats were 217 ± 88.5 for the two-stage sham operated rats and 208 ± 19.6 for the saline injected rats[27]. The differences were not significant from each other. Hence the behavior of these further control rats shows that the trauma from surgery (for all of the groups of rats) "had little, if any, bearing on subsequent learning" (Patrissi and Stein, 1975, pg. 477).

Thus the basic idea is that the brain can recover from fairly significant lesions to recover to a normal level of performance, assuming that the damage occurs over a suitable period of time. If the damage afflicting a particular area is sudden and occurs in a one-time fashion, the deficits are significantly worse than if the damage to the same area is gradual. More particularly, if the damage occurs gradually, the brain can find ways to implement the same psychological functions using whatever is left.

7.2.2 Evaluating the Case Study's Evidence for Multiple Realization

[27] Notice also that these numbers are higher than both the 20-day and 30-day seriatim rats (which, of course, had neural tissue removed and not just a 2-stage sham surgery), further demonstrating the psychological similarity of the seriatim rats (who had sufficient time to recover) and the sham surgery rats.

For the present purposes, this experiment is interesting because it shows that the same psychological ability can be multiply realized in different neurological kinds. In order to see this, it is best to start by analyzing the psychological ability. The authors never make it clear what spatial alternation tasks are used to measure, but generally these tasks show the extent to which a subject is capable of suppressing previously correct responses. The alternation of the reward forces the organism to suppress the information it has of its former location. The selection of the frontal cortex for resection makes sense, too, in the context of this task. This area in the rat (or at least sub-areas of it), much like the human frontal cortex, is involved (among other things) in planning actions for the future. Part of the process of planning for the future involves forming representations of different courses of actions, calculating their consequences, comparing those consequences to the goals and preferences of the organism, and selecting one option while also *suppressing* the other alternatives. This is how the cognitive sciences conceive of our decision-making capacities. The neural implementation of these abilities is, roughly speaking, the frontal lobe. Thus, in the same way that ventral frontal patients in humans show dysfunctions in decision-making (Bechara et al, 2000), so too the one-stage rats are very impaired in their ability to make decisions that will lead them to obtaining the rewards and reaching criterion. The psychological ability in question is something like the ability to suppress particular choices when making decisions—viz., in this case, to inhibit previously correct responses (from biasing the behavior of the rat to choose the wrong arm). The lack of a significant difference in the controls, the two-stage, 30-day interval rats, and the two-stage, 20-day interval rats suggests that these animals are psychologically homogeneous in at least this respect.

What about the neurological kinds? Here the answers are less clear, but enough to establish the multiple realization of the ability to inhibit behavioral responses. What is known for sure and worth mentioning is that the authors took great care to ensure that the rats in all of the lesioned groups received as similar damage as possible. They performed a one-way analysis of variance to determine whether the different groups had sustained significantly different damage. The outcome of the analysis was that the "resultant F value was not significant . . . indicating that the amount of frontal tissue removed was approximately the same across the groups with different interoperative intervals". (Patrissi and Stein, 1975, pg. 476) So there is no worry about differences in the extent of the resections explaining why some rats suffered from greater deficits than others.

But the key question obviously is: in which new, distinct Brodmann areas are the relevant psychological functions realized? Unfortunately, it is impossible to tell. The types of functional brain scans available since the early 90s—so popular for localizing psychological functions—were unavailable during these experiments. The researchers could have used something like single-cell recordings to locate the areas in all of the groups which responded differentially according to the task, but the point of their research was not to locate the neurological substrates of response inhibition (or anything else), so it is impossible to tell exactly where the psychological functions have migrated in the lesioned rats. In neurologically normal rats, the prelimbic cortex (see Delatour and Gisquet-Verrier, 1996; Jonkman et al, 2009) and the infralimbic cortex (see Chudasama and Robbins, 2003) of the frontal cortex have been implicated in the inhibition of Pavlovian and instrumental responses. The infralimbic cortex is located in Brodmann area 25, while the prelimbic cortex is in Brodmann area 32. These areas do not participate in the learning of associations, but simply aid in the suppression of responses that are inappropriate for the situation. The researchers removed these areas in the lesioned rats, so they obviously cannot be instantiating the psychological abilities being tested. But it is unfortunately anyone's guess which neural regions are performing these executive tasks in the experimental rats.

What is clear, keeping in mind that the damage was fairly constant, is that there is little psychological difference in the first three groups, though an entire lobe is missing in the seriatim lesion rats. As the one-stage lesion rats show, the frontal cortex is important for reaching criterion. If suppressing previously correct responses were mediated by, say, the temporal lobe, then all of the rats would have performed the same. The obvious explanation is that the two-stage lesion rats must have been using some entirely different lobe, thus some entirely different neurological kind, to reach criterion. I obviously do not know which areas those are; the authors include no information about what areas are now processing the relevant information. But it also seems unnecessary to know. The prelimbic and infralimbic cortices are part of the frontal cortex of the rat, leaving little doubt about their complete destruction since the researchers removed the entire frontal lobe in the experimentals. Setting aside dualism, some other genuinely distinct neural area(s) must be carrying out the inhibitory processing. The rats thus provide rather stark evidence in favor of multiple realization. Here, within an individual, there is significant psychological similarity realized in different neural structures.

Patrissi's and Stein's rats also do not run afoul of any taxonomic grain considerations. When the grain sizes for both the psychological and neurological sides are held equal, excepting perhaps the most implausibly coarse grain, the results still show that the psychological kinds are multiply realized in the neurological ones. To explain, remember that there are only four lobes in the rat brain. In the present case, the psychological grain is somewhat fine. Of all the executive tasks to take into consideration (integrating incoming sensory information, exploitation of top-down connections (to modulate attention), evaluation of the importance of stimuli and behavioral responses, etc.), the suppression of inappropriate responses is the only psychological function the experimenters are directly testing. On the other hand, the classificatory grain for the neurological kinds is about as coarse as possible. The authors did not simply remove the medial wall of the prefrontal cortex (the location of the infralimbic and prelimbic cortices), but the entire frontal cortex. There is an obvious mismatch of the grain sizes, but it is not one that runs afoul of the spirit of Bechtel's and Mundale's principle. Their suggestion is that matching a coarse-grained psychological kind with a fine-grained set of neurological kinds provides illusory evidence for multiple realization, not a matching of fine-grained psychological kinds with coarse-grained neurological kinds. If anything, as we have noted above, the latter kind of mismatch provides *better* evidence for multiple realization than contexts where the grain size is matched. If some rather fine-grained psychological kind (e.g., inhibition of behavioral responses) can be shown to be multiply realized even in very coarse-grained neurological kinds (e.g., those typed by reference to entire lobes), then a fortiori that psychological kind must be multiply realized in neurological kinds typed at all other finer levels, too.

So even if one implausibly says the entire frontal lobe is necessary for reaching criterion (adopting a coarse-grained view of neurological kinds), its entire removal in the lesioned subjects still shows that some other (area of another) lobe is carrying out the processing, providing evidence that some distinct neurological kinds are processing the same information the frontal lobe formerly was. If it is necessary to match that coarse level of grain on the psychological side, the case is a cinch. At that level, it is arguably true that *all* of the rats are the same (and not just the sham, 30-day, and 20 day interval experimentals). After all, even the one-stage lesion rats reached criterion. The increased number of days it took that group (relative to the others) could be overlooked if what constitutes psychological similarity is as loosely construed as the neurological case (assuming our ignorance of the specifics forces us to retreat to that very coarse-grained perspective).

7.2.3. The GRP More Generally

These serial lesion cases are specific instances of the more general GRP, which holds, to repeat, that the brain will reorganize neurological processes in response to brain damage of all kinds if it is allowed enough time to adapt. This is indicated by the Patrissi and Stein study where it they demonstrate that a small interoperative interval did not result in as much functional sparing as with the patients with longer intervals between the lesions. But this effect is hardly peculiar to the surgeon's scalpel. For instance, it is well known that damage to the pars compacta of the substantia nigra results in movement disorders like Parkinson's. The pars compacta is an important area for the production of the neurotransmitter dopamine. One of its most important projection zones is the striatum, which is critical for producing movements. Once this area of the substantia nigra begins to deteriorate, dopamine transmission in the striatum lowers and Parkinson's is the result. It is also well known that the substantia nigra as a whole deteriorates in all humans, much the same as humans naturally lose brain tissue of all kinds over the course of their lives. What is remarkable about the loss of the pars compacta, however, is that humans are able to overcome around 80% of its destruction without any deficits (see Nass et al, 2008 and Bernheimer et al, 1973). Most people typically have enough this area to last a lifetime (even through its natural deterioration), but through other genetic factors (or through an amphetamine addiction) the rate of its deterioration can be accelerated to the point that the 80% threshold is broken. From the perspective of multiple realization, the interesting fact about the substantia nigra is that it shows again the ability of the brain to survive functionally while its tissue is damaged—provided that the damage occurs slowly enough to allow all of the necessary rewiring.

Another example of the GRP is the difference between slow-growing tumors (e.g., low grade gliomas) and acute strokes. In the same way that serial lesions typically offer much better chances of functional recovery, so too slow-growing tumors are usually accompanied by little or no deficits—in stark contrast with the comparatively debilitating results of strokes. In fact, in about 80% of patients, the functional reorganization is so substantial and quick that the presence of low grade gliomas is not revealed by deficits in behavior but usually by the sudden occurrence of seizures (DeAngelis, 2001). One study (Duffau et al, 2003) accumulated data over 5 years on 103 subjects with cortical or subcortical low grade gliomas who suffered from only mild deficits or no deficits at all. After the tumors were taken out, which also meant removing the cortical areas responsible for the psychological functions, 94% of the patients recovered to preoperative psychological levels within 3 months. Thus, the same general kind of principle in the serial lesion cases holds true in the literature on tumors and strokes: as long as the brain has time to rewire and find other ways to carry out the same processes, it can usually overcome damage. As neurosurgery has improved, it has become possible to remove low-grade gliomas before they become so large as to affect behavioral performance. As a result, it has been found that fairly major regions of the brain can be removed without creating measurable behavioral problems, including Broca's area, the left and right insula, the sensorimotor cortex, the supplementary motor area, the left posterior parietal cortex and even the temporal lobes (see Desmurget et al, 2007, pg. 899, for citations of these examples). Hence the GRP one finds at work in the literature on serial lesions is mirrored by findings in neurooncology. Individuals who are psychologically the same can nevertheless suffer from profound destruction of the brain, indicating that the same psychological kinds can be instantiated in many different neurological kinds.

7.2.4 Conclusion

It is surprising that the philosophical work on multiple realization has never contained a discussion of anything like the GRP or, more specifically, the serial lesion effect. The literature on the effect is quite old, too, with the Patrissi and Stein paper cited here coming comparatively late in the game. Perhaps this owes to the fact that information in the neurosciences can be obsolete within months (much less decades) and it is usually not a good idea use studies of that kind as the cornerstone of an argument. But the lack of current interest in the serial lesion effect has nothing to do with its falsification or being out of date. Interest in the subject waned simply because researchers realized the effect was quite real; there is no pressing reason (or funding) for replicating an oft-replicated experimental effect. Instead, the mechanisms underlying the functional sparing rightly came to dominate research. The final paragraphs of Patrissi's and Stein's study, in fact, offer a brief discussion of the possible mechanisms of recovery for their frontal lesion rats. So perhaps there is reason for wonder why cases like Patrissi's and Stein's have never made their way into the discussion on multiple realization and identity theory, but none for caution about the datedness of the evidence they provide.

Regardless of its lack of history in the literature, however, the serial lesion effect and the GRP more generally have a considerable impact on philosophical positions like the identity theory and multiple realization. These cases show that the type of counterfactual dependence of psychological kinds on neurological kinds is not what the identity theorist's position requires. An identity of the one kind with the other would presumably have consequences like, 'if the neurological kind had been destroyed, the psychological kind would have been destroyed, too.' The GRP, however, demonstrates that these types of counterfactuals are, strictly speaking, not true. Rather, if the neurological kind had been destroyed *in a sudden fashion*, then the psychological kind would have been similarly destroyed. On the other hand, if the neurological kind were destroyed in a piecemeal, gradual manner, then the psychological kind would have been spared. Such a refinement of the counterfactuals is not merely pedantic. A genuine identity of psychological kinds with neurological kinds implies the former, empirically denied counterfactual. Thus the updated iteration of this counterfactual, which takes into account the slow-going or sudden nature of the damage, shows that psychological kinds are simply not identical with neurological kinds.

Additionally, the GRP shows something important about lesion studies and using them as evidence for philosophical positions. Of course it is philosophically interesting (primarily for ontological questions about dualism, multiple realization, and identity theory, but presumably also for questions about consciousness) that the removal of some area of the brain has psychological consequences, but the GRP shows there is another dimension to the lesion literature. Just as important as *where* one digs around in the brain, it is important to pay attention to *when*, and over how much time, that digging occurs. The GRP shows that the temporal nature of the damage is a critical factor in determining the psychological consequences of brain damage. Thus the temporal nature of the damage is a critical factor in determining the philosophical consequences of the damage as well.

Section 7.3: The Famous Cross-Wired Ferrets Case

Last is the case of the rewired ferrets. Their fame (or infamy) in the literature as failures of evidence for multiple realization has always struck me as undeserved. Here I would like to review briefly the findings of Sharma et al (2000), then discuss some of the remarks made about the case by Shapiro and Polger. They are responsible for the paper's recognition in philosophy and provide the best criticism of the results as they apply to multiple realization. I think, in light of the other cases we have discussed so far, I can meet their criticism and restore the usefulness of their findings as evidence for multiple realization. So the final case here will be much less about methods, experimental models and how well they support the authors' conclusions, and more about exploring an empirical work more deeply in a philosophical context.

7.3.1. The experiment on the ferrets: why the researchers rewired and what they found

But I do not want to skip entirely any exegesis of Sharma's, Angelucci's, and Sur's paper. We have to make sure of the empirical content before we can responsibly discuss its philosophical implications, so let me devote a page or two to an explanation of what they were looking for, what they found, and how they interpreted the significance of what they found.

Their overall goal, unsurprisingly, was not to resolve any dispute in philosophy. They were not even seriously interested in the question about how extensive plastic changes can be. Rather their main focus was to investigate the degree to which afferent activity shapes thalamocortical and intracortical connections. That is, they wanted to know how important incoming sensory input is for developing the connections between the thalamus and the cortex and also connections between cortical areas. One might expect, for instance, that most of the structures of the brain—particularly cortical areas that are highly specialized for a particular modality—are, as it were, pre-programmed, so that incoming neural activity, while using the scaffolding provided innately, never has much of an influence in building up the cortical configuration. Their focus was mainly on the presence of orientation columns and horizontal connections which one typically finds in the primary visual cortex (V1). They cite evidence that, for instance, whole maps of orientation columns rely only on an intrinsic organization of connections that are not driven by afferent activity. Thus their overall goal is to rearrange connections within the brain and see what effects afferent activity has in other areas which typically never receive that information. This is why they are interested in rerouting visual information into the primary auditory cortex (A1)—to hope that this area of the cortex comes to resemble V1, indicating that sensory input has a large effect in constructing the cytoarchitectural design of a piece of cortex.

In order to reroute information from the retina, they ablated Brodmann areas 17, 18 (areas in the occipital lobe, in the heart of the usual visual processing areas), the superior colliculus (another target of retinal input, used to direct eye movements), and the inferior colliculus (an area with major input into A1) in neonate ferrets.[28] One result is that the thalamic area (the LGN) which relays retinal information to areas 17 and 18 atrophies fairly seriously. In response to this thalamic decay, combined with the loss of input from the inferior colliculus, the thalamic area (the MGN) which projects to the main auditory processing areas (A1 and the auditory

[28] For more details of the procedure, see Sur et al (1988).

fields) receives projections from the retina. Ultimately, this causes visual information to arrive in the main auditory processing areas, giving the authors a golden chance to witness the effects of sensory input on the organization of the cortex.

What they find, of course, is that rewired A1 comes to appear organizationally much more like V1 than in ferrets with a normal A1. For the first obvious difference, many of the neurons in the rewired A1 are visually sensitive. They provide evidence about orientation tuning of certain cells in rewired A1, so there are cells in the rewired cortex that respond to visual rather than auditory stimuli. But they also found evidence more germane to their goal of showing the degree of activity-dependent plasticity. There was sharp orientation sensitivity found in cells of rewired A1, with tuning that was similar to that in normal V1. Orientation sensitivity refers to the responsiveness of single neurons to a particular orientation of a bar of light at an angle (straight up and down, diagonal, etc.). Some cells in V1 will only fire when such a bar at just the right angle crosses into its visual receptive field, while remaining comparatively silent when anything else appears in the field. Cells in the rewired A1 showed the same kind of tuning. Next, there were entire orientation maps within rewired A1 that, though less orderly than what is found in V1, were systematically arranged. Orientation maps refer to the organization of these cells that respond only to a particularly oriented stimuli. In V1, all of the cells which respond to the same stimuli are grouped in an orderly manner together. Cells in rewired A1 also exhibited, to a lesser extent, this organization. Finally, long-range horizontal connections within rewired A1 resemble much more closely the connections found in V1 rather than normal A1. Hence the overall architecture of the major connections of the functionally similar cells in rewired A1 appears somewhat more like V1 than normal A1.

From these results, Sharma et al conclude that sensory input goes a long way towards determining the functional organization of the cortex. Aside from being interesting for Shapiro (who would be impressed by the fact that higher-level visual organization is dependent on such lower-level organization found in the orientation sensitivity and ocular dominance columns), this conclusion is not the one philosophers of mind probably care much about. Instead, the important philosophical implication of this experiment is that we have the auditory cortex processing visual information—what, at least on the surface, is a promising case for the multiple realization of vision (at least in ferrets). Nevertheless, it is worth paying attention to the details of the experiment, as we will see in a moment.

7.3.2. Evaluating the evidence the ferrets provide for multiple realization: what Shapiro and Polger have to say and my response

What exactly has been the reaction of those like Shapiro and Polger? Both agree, for largely the same reasons, that the case of the rewired ferrets does not provide any evidence for multiple realization. To provide a few quotes, Shapiro says: "[The experiment] affords neither a case in which the *same* function is realized in two different ways, nor a case in which the same function is realized in two *different* ways" (Shapiro, 2004, pg. 64; author's emphasis). Polger, echoing and extending Shapiro, says:

"because the rewired ferrets have significant visual deficits vis-à-vis normal ferrets . . .[and] because the case is better described as one in which visual cortex is "grown" in or "moved" to the location normally occupied by auditory cortex . . . [and] because difference in gross brain location is irrelevant to the function of the system in vision . . . [and] because the normal and rewired visual cortexes [sic] do their jobs in much the same way (insofar as they do the "same" thing at all) . . . the rewired ferrets are not good evidence of multiple realization" (Polger, 2009, pg. 468)

So all of the lines one would anticipate the enemy of multiple realization to take have been advanced. The extent to which normal and rewired ferrets share the "same" visual abilities, whether normal V1 and rewired A1 should be considered different lobes, and the way in which the information is processed—all of it is scrutinized for the same reasons I have been discussing for the entire dissertation. I take each worry in order.

First we have the main problem for the ferrets as evidence of multiple realization: the dissimilarity of their visual abilities. Both Shapiro and Polger complain that the vision of the rewired ferrets is poorer than in normals, as measured by the experimentals' lesser ability to "detect gratings at a higher frequency (a greater spatial resolution)" (Shapiro, 2004, pg. 64). The point, of course, is that multiple realization requires the same psychological kind to be realized in relevantly different neurological kinds. If the rewired ferrets' A1 does not realize the same visual ability, then they cannot provide any evidence for multiple realization. There are two problems, I think, with this objection—one I will mention now, one I will save for later.

The first problem is that the quote Shapiro provides about the ferrets' lesser visual ability is from a different paper than Sharma et al. He cites von Melchner, Pallas, and Sur (2000), another paper in *Nature* investigating how the modality of a piece of cortex is at least partially determined by extrinsic inputs from the sensory organs. The case I have been considering, which Shapiro also cites (and appears to be referring to when dismissing its significance as evidence for multiple realization), mentions nothing about testing the rewired ferrets for their visual acuity. Their only goal was to demonstrate that afferent activity plays a structuring role for the organization of the cortex, not to show that any similarity of acuity held between the ferrets.

This point is complicated further by an ambiguity about how retinal input was redirected to A1 for the ferrets in each experiment. Sharma et al cite an older paper (Sur, et al, 1988), without giving any further information about how they directed retinal axons to the MGN. Sur and colleagues are, on the other hand, very explicit about their procedure: they ablate areas 17 and 18, the superior colliculus, and the inferior colliculus, causing the retina to project to the MGN (rather than the LGN only). Melchner, Pallas, and Sur are fairly clear about the procedure they used (also providing a figure): they ablated the brachium of the inferior colliculus and the superior colliculus, eliminating important contralateral and ipsilateral retinal projections which ultimately cause the optic tract to connect with the MGN (and hence on to A1). But it is not clear, for instance, whether they also cleared out areas 17 and 18. I have no idea whether those extra resections would make a significant difference in the visual acuity of rewired A1, and it is entirely likely that Melchner et al performed identical procedures to induce visual activity in A1 of their ferrets without making it explicit (since these authors, together and with others, have published programmatic research mostly having to do with effects of cross-wiring brains in neonate mammals for many years). Also, Sharma and colleagues do make it clear that the cortical organization of rewired A1 looks like a raggedy copy of normal V1, which offers some reason for thinking that the corresponding visual ability of the ferrets will be raggedy, as well. My point is only that it is not exactly fair to pin a particular result on one experiment based on the research of another, particularly when the former was not interested in the conclusions of the latter, and definitely when it is not clear if the methods the latter used to obtain the result are different from the methods used by the former. Accordingly, it is not unreasonable to wonder if the rewired ferrets of Sharma et al have better visual acuity than those from Melchner et al, or if they have any visual deficits at all. We just do not know. Notice, to be fair, that I have the opposite problem: in order to show that the ferrets offer evidence in favor of multiple realization, I need to establish a psychological similarity between rewired and normal ferrets. When I tackle what I take to be the other problem for Shapiro's (and Polger's) criticism, I will discuss this further problem for me.

Next is the discussion of whether the rewired A1 is genuinely different from normal V1. This is important, obviously, for establishing that the visual abilities are *multiply* realized. If these areas of the cortex are not suitably distinct, then the vision gained by the rewired ferrets does not present a case of vision being multiply realized—rather it would only confirm that vision is realized in something like what one finds in V1. I want to break this end of the criticism into two parts. The worry about how rewired A1 develops similar organizational attributes to normal V1 raises two different possibilities, both of which I respond to separately: that, essentially, the primary visual cortex has simply 'grown' in a different location, *and* that the rewired A1 functions (to the extent that it realizes vision suitably like that found in V1) the same as V1 (hence, according to Shapiro's criterion for genuinely distinct realizations, does not constitute a multiple realization).

I will begin with the second worry. Since I have said so much already about Shapiro's criterion, there is not much to add here. Besides presenting a real, empirical case where Shapiro's and my respective criteria disagree, the ferrets will not drag us over any new ground. I think it is fairly uninteresting that rewired A1 appears to process whatever visual information it does in the same way as normal V1. I take it, if anything, that similar processing shows that we do have something like the same psychological process instantiated—and, as I have admitted above, it might even be the case that these cortical organizations are something like universal constraints on the visual acuity found in most mammals, or at least ferrets. From my perspective, what matters for distinguishing neurological states is the taxonomy of the science that governs the brain. In this case, there are clearly two different kinds, even from a rather coarse grain: the visual cortex and the auditory cortex, each of which are found in entirely different lobes in the cortex. According to a Brodmann taxonomy—which I have suggested as a suitably non-psychologically-driven neurological taxonomy—the primary visual cortex is found in area 17, while the primary auditory cortex is essentially areas 41 and 42. Hence I do not buy Shapiro's criticism that the rewired ferrets do not demonstrate a difference in realizer kinds. It might well be that rewired ferrets do not share the same visual abilities as normal ferrets, but, according to neurological classifications, what visual acuity they do possess is clearly realized in a distinct substrate.

Next is Polger's attempt to show rewired A1 is not a different realizer of the ferret's visual abilities (than V1). Instead of focusing on the nature of the processing, he suggests that "the case is better described as one in which visual cortex is "grown" or "moved" to the location normally occupied by auditory cortex" (Polger, 2009, pg. 468). This point stems, I suppose, from the idea that the modality of the cortex is not necessarily hard-wired, but is subject to activity upstream. Sharma and colleagues, of course, were ultimately trying to prove this about the organization of the cortex. It is not crazy to think that the success of the rewired ferrets is enough to say that the visual cortex has 'grown into' another location—but I do not buy it.[29] There is still the counterpoint, which is acknowledged in the *Nature* article, that many features of the cortex are simply 'hard-wired'. As Sharma et al say, "[t]he differences between orientation maps and horizontal connections in rewired A1 and V1 suggest constraints on activity-dependent

[29] Even if I were forced to buy it, though, that would indicate that neurological kinds are typed with reference to psychological kinds, which, as I have pointed out repeatedly, dooms the case for the identity theory. Then the problem wouldn't be psychology being autonomous from neuroscience, but the other way around. See below for yet more elaboration on this point.

plasticity". (op cit, pg. 846) Also, the very fact that Brodmann has labeled the areas differently shows that there are cytoarchitectural differences between what is typically A1 and V1—enough difference, I take it, to establish them as genuinely different neurological kinds. Anyway, if that is not enough, Melchner et al take a dimmer view of the similarity that obtains between rewired A1 and V1, saying that "the cortico-cortical connections of rewired A1 remain essentially the same as in normal animals". (op cit, pg. 875) This finding contradicts the evidence Sharma et al produced, showing that long-range horizontal connections within rewired A1 were similar to those found in normal V1. Hence a more plausible description of the rewiring is that the primary auditory cortex has taken on some of the organization of the primary visual cortex, while retaining many of its normal structural features, *not* that V1 has jumped over to the temporal lobe.

That the rewired A1 comes to process visual information, even roughly using the same sort of fine-grained algorithm (indicated by the presence of orientation selective cells, for instance), is also not a strong enough reason to conclude that V1 has moved into a new location in the cortex. Again, I think this move is a bad one for the same reason I mentioned in the chapter about how to count brain state types. If our neurological taxonomy is too psychologically guided—to say nothing about the odd consequences with silicone chip brains—then we trivialize the identity theory; it becomes the product of typing kinds of one theory in a one-to-one correlation with kinds from another theory, not an independent discovery that the kinds picked out at the one level line up in a one-to-one fashion with the kinds of another level. So we definitely do not want to say that rewired A1 taking over the left hemisphere's visual processing by that very fact turns it into the 'new' V1. Polger's insistence that V1 has simply 'grown into' this other cortical location comes very near, I think, to making just this mistake.

We need to look at the cortex from a non-functionally-minded perspective in order to evaluate multiple realization properly. When we do—as the differences in Brodmann areas show—we get distinct neurological kinds. So, in the end, I do not find anything too compelling about either Shapiro's or Polger's challenge to the idea that rewired A1 is a bona fide different neurological kind than V1. If rerouting vision, even a subdued form of it, to an entirely different lobe is not sufficient for establishing a really distinct realizer, then I am not sure if there could be any evidence for multiple realization.

Finally, I want to return to the worry about similarity of psychological kinds (I had said there were two problems with this attack, one of which I would bring up later). This is a worry in the present case because the rewired ferrets, at least in another similar experiment to the one we have been worrying about, suffered from visual deficits from which normal ferrets did not. A fairly appealing line of response to this worry, I think, starts from Bechtel's and Mundale's point about grain size. From a suitably coarse level of grain, the case of the rewired ferrets presents good evidence for multiple realization. That is, from a coarse-grained perspective, the vision achieved by the rewired ferrets is similar enough to that found in normals; and certainly there are differences in the brain state kinds that realize that vision—from a fine-grained level and right on down to very coarse-grained levels. Even with very blurry vision, it is hard not to be able to distinguish the occipital from the temporal lobe. So, avoiding an equivocation on grain by remaining rather coarse about how to classify the kinds in question, the rewired ferrets do provide decent evidence for the claim that psychological kinds are multiply realized in neurological kinds. Though I do not think that putting the matter this way really knocks the identity theory out, I do think this line is persuasive and hints at what I take to be the real take-home point from the ferrets.

7.3.3. Conclusion

That point is that we really have two competing hypotheses which can be judged by the rewired ferrets. On the one hand, we have the hypothesis that psychological kinds are identical to neurological kinds, and, on the other, the hypothesis that the same psychological kinds are realized in many different neurological kinds. Straightforwardly, the ferrets do not support the former hypothesis. But that is not too damaging. What really sinks the view is that a prediction one would likely make on the basis of that hypothesis is that the elimination of Brodmann areas 17 and 18 would render the ferrets completely blind. If the kind of vision found in rewired A1 were really identical to those occipital areas, then the (all-at-once) elimination of the neural substrates ought to cause the disappearance of that vision, too. Yet it is clear the rewired ferrets have achieved some sort of diminished form of vision despite the ablation of the all-important cortical areas. The only way to save the hypothesis in the face of this empirical fact, I think, is to argue that rewired A1 is just another token of the V1 kind. From the standpoint of neurological taxonomy, however, A1 and V1 are flatly different cortical kinds. So the ferrets do appear rather problematic for the view that psychological kinds are in register with neurological ones.

Of course this has not stopped people defenders of the identity theory like Polger from trying to argue that the rewired A1 is just the new V1. It is a move that is available to the identity theorist which saves it from the apparent counterevidence. But such a move could save the view at any turn. An identity theorist could use the same tactic against the other two cases of multiple realization I have produced. The question is whether the move carries any independent plausibility. It is obviously not sufficient just to say anything that will save the identity theory.

Polger seems to have got a lot of mileage out of the identity theory simply by not coming to grips with what a brain state type is. If one chooses not to be explicit about these types, then it allows room to be slippery about how the evidence is supposed to be interpreted. In the ferrets' case, the slipperiness allows him to maintain that the rewired A1 is just where the primary visual cortex has 'grown'. Having no real philosophical constraints makes this possible. But fortunately there are empirical constraints that can trump Polger's ad hoc moves. Even if one does not like my proposed taxonomy of Brodmann areas, there is no other way recognized by anyone working in the neurosciences of counting the neurological kinds without drawing a sharp difference between these two areas. In itself this fact is enough to pass over the worry about whether the rewired A1 is genuinely distinct from normal V1.

Do the ferrets support the other hypothesis? One must admit that it is not *obvious* they do. Unless we adopt a rather coarse-grained view of psychological (and neurological) kinds, we do not have any great reason for saying that the vision of the rewired ferrets is of the same type as that found in normal ferrets. Nevertheless, it is clear that a prediction one would make on the basis of the latter hypothesis is that (at least roughly) the same psychological kind would be correlated with activity in different types of neurological states. And it is here that the ferrets come the closest to supporting either hypothesis. Plainly something like the same visual ability is realized in vastly different areas of the cortex. It is not an absolute cinch, but the multiple realization hypothesis certainly comes out looking better than the identity theory hypothesis. This is worth mentioning when considering the case of the rewired ferrets.

From an information processing standpoint, the ferrets also offer excellent evidence that what is important psychologically is the ability to capture and process information, not necessarily to have a particular kind of realizer. So long as there are neurons to encode the right information, even the primary auditory cortex is capable of realizing vision. Again, this seems to support further the idea that psychological processes can be realized in many different substrates. It is even possible that the surgical intervention required to rewire the cortex resulted in scar tissue which disrupted the processing of information, or that the cytoarchitectural profile of the primary auditory cortex makes it unsuitable for processing the necessary information optimally, making the achievement of even a limited form of vision (perhaps) quite a striking demonstration of how inessential any particular kind of the cortex is for realizing psychological kinds and processes.

Bibliography

Ades and Raab (1946). "Recovery of motor function after two-stage extirpation of area 4 in monkeys". *Journal of Neurophysiology* (9): pp. 55 – 60.

Aizawa (2009). "Neuroscience and Multiple Realization: A Reply to Bechtel and Mundale". Synthese (167): pp. 493 – 510.

Bechara, Tranel, and Damasio (2000). "Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions". *Brain* (123): pp. 2189 - 2202.

Bechtel and McCauley (1999). "Heuristic Identity Theory (or Back to the Future): the Mind-Body Problem" Against the Background of Research Strategies in Cognitive Neuroscience." *Proceedings of the 21st Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates.

Bechtel and Mundale (1999). "Multiple Realizability Revisited: Linking Cognitive and Neural States." *Philosophy of Science* (66): pp. 175-207.

Bernheimer, Birkmayer, Hornykiewicz, Jelliner, and Seitelberger (1973). "Brain dopamine and the syndromes of Parkinson and Huntingon. Clinical, morphological, and neurochemical correlations". *Journal of the Neurological Sciences* (20): pp. 415 – 455.

Bickle (2003**).** *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Boston: Kluwer Academic Publishers.

Block (1995**).** "A Confusion about a Function of Consciousness'. *Behavioral and Brain Sciences* (18): pp. 227 – 247.

Block and Fodor (1972). "What Psychological States are Not". *Philosophical Review* (81): pp. 159 – 189.

Boyd (1999**).** "Kinds, Complexity and Multiple Realization," *Philosophical Studies* (95): pp. 67-98.

Brodmann (1909**).** *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Barth, Leipzig. Translated into English and edited by L.J. Garey, 1994: Brodmann's 'Localisation in the Cerebral Cortex', Smith-Gordon, London.

Chalmers (1996**).** *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.

Chudasama and Robbins (2003). "Dissociable contributions of the orbitofrontal and infralimbic cortex to pavlovian autoshaping and discrimination reversal learning: Further evidence for the functional heterogeneity of the rodent frontal cortex". *Journal of Neuroscience* (23): pp. 8771 – 8780.

Cohen, Forman, Braver, Casey, Servan-Schreiber, and Noll (1997). "Activation of the Prefrontal Cortex in a Nonspatial Working Memory Task with Functional MRI". *Human Brain Mapping* (1): pp. 293 – 304.

Courtney, Ungerleider, Keil, and Haxby (1997). "Transient and sustained activity in a distributed neural system for human working memory". *Nature* (386): pp. 608 – 611.

Courtney, Petit, Maisog, Ungerleider, & Haxby (1998). "An area specialized for spatial working memory in human frontal cortex". *Science* (279): pp. 1347 - 1351.

Cowey (1964). "Projection of the Retina on the Striate and Prestiate Cortex of the Squirrel Monkey Saimiri Sciureus". *Journal of Neurophysiology* (27): pp. 366-393.

DeAngelis (2001). "Brain Tumors". *New England Journal of Medicine* (344): pp. 114 – 123.

Delatour and Gisquet-Verrier (1996). "Prelimbic cortex specific lesions disrupt delayed-variable response tasks in the rat". *Behavioral Neuroscience* (110): pp. 1282 – 1298.

Dennett (1990). "Quining Qualia". in A. Marcel and E. Bisiach, eds, *Consciousness in Modern Science*. New York: Oxford University Press.

Desmurget, Bonnetblanc, and Duffau (2007). "Contrasting acute and slow-growing lesions: a new door to brain plasticity". *Brain* (130): pp. 898 – 914.

Duffau, Denvil., and Capelle (2002). "Long term reshaping of language, sensory, and motor maps after gliomas resection: a new parameter to integrate in the surgical strategy". *Journal of Neurology, Neurosurgery, and Psychiatry* (72): pp. 511 – 516.

Duffau, Capelle, Denvil, Sichez, Gatignol, Taillandier, Lopes, Mitchel, Roche, Muller, Bitar, Sichez, and van Effenterre (2003). "Usefulness of intraoperative electrical subcortical mapping during surgery for low-grade gliomas located within eloquent brain regions: functional results in a consecutive series of 103 patients". *Journal of Neurosurgery* (98): pp. 764 – 778.

Felleman and Van Essen (1991). "Distributed hierarchical processing in the primate cerebral cortex". *Cerebral Cortex* (1): pp. 1-47.

Finger, Walbran, and Stein (1973). "Brain damage and behavioral recovery: serial lesion phenomena" *Brain Research* (63): pp. 1 – 18.

Finger (1994). *The Origins of Neuroscience*. New York: Oxford University Press.

Flourens (1824). “Investigations of the properties and the functions of the various parts which compose the cerebral mass”, reprinted in G. von Bonin (transl) *Some Papers on the Cerebral Cortex*. Thomas: Springfield Illinois: pp. 3 – 21.

Fodor (1968). *Psychological Explanation: An Introduction to the Philosophy of Psychology*. New York: Random House.

Fodor (1974). “Special Sciences (or: The Disunity of Science as a Working Hypothesis)”. *Synthese* (28): pp. 77 – 115.

Fodor (1975). *The Language of Thought*. New York: Thomas Crowell.

Fodor (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Fodor (2000). *The Mind Doesn’t Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.

Fodor (2008). *LOT 2: The Language of Thought Revisited*. New York: Oxford University Press.

Freeman, Riesenhuber, Poggio, & Miller (2001). “Categorical representation of visual stimuli in the primate prefrontal cortex”. *Science* (291): 312-316.

Funkhouser (2007) “A Liberal Conception of Multiple Realizability”. *Philosophical Studies* (132): pp. 467 – 494.

Goldman-Rakic (1995). “Cellular basis of working memory”. *Neuron* (14): pp. 477 - 485.

Henschen (1893). "On the Visual Path and Centre", *Brain* (16): pp. 170-180.

Hubel and Wiesel (1962). "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex". *Journal of Physiology* (London) (160): pp. 106 - 154.

Hubel and Wiesel (1965). "Receptive Fields and Functional Architecture in Two Non-striate Visual Areas (18 and 19) of the Cat". *Journal of Neurophysiology* (28): pp. 229 - 289.

Jacob (2008). “What do mirror neurons contribute to human social cognition?”. *Mind and Language*. 23 (2): pp. 190 – 223.

Jonkman, Mar, Dickinson, Robbins, Everitt (2009). “The rat prelimbic cortex mediates inhibitory response control but not the consolidation of instrumental learning”. *Behavioral Neuroscience* (123): pp. 875 -885.

Kennard (1938). “Reogranization of motor function in the cerebral cortex of monkeys deprived of motor premotor areas in infancy”. *Journal of Neurophysiology* (1): pp. 477 – 496.

Kennard (1942). “Cortical reorganization of motor function” *Archives of Neurology and Psychiatry* (48): pp. 227 – 240.

Kim (1992). “Mulitple Realization and the Metaphysics of Reduction”. *Philosophy and Phenomenological Research* (52): pp. 1 – 26.

Kim (2005). *Physicalism, or Something Near Enough*. Princeton: Princeton University Press.

Kim, S (2002). “Testing Multiple Realizability: A Discussion of Bechtel and Mundale”. *Philosophy of Science* (69): pp. 606 – 610.

Kleitman and Camille (1932). “Studies on the physiology of sleep. VI. The behavior of decorticated dogs”. *American Journal of Physiology* (100): pp. 474 – 480.

Lashley and Clark (1946). “The cytoarchitecture of the cerebral cortex of Ateles: A critical examination of architectonic studies”. *Journal of Comparative Neurology* (85): pp. 223–305.

Lewis (1972). “Psychophysical and Theoretical Identifications,” *Australasian Journal of Philosophy*, 50: 249–58.

Luna, Thuslborn, Strojwas, McCurtain, Berman, Genovese, and Sweeney (1998). “Dorsal cortical regions subserving visually guided saccades in humans: an fMRI study”. *Cerebral Cortex* (8): 40 – 47.

Marslen-Wilson (1987). “Functional Parallelism in Spoken Word Recognition”, *Cognition* (25): pp. 71 – 102.

Millikan (1999). “Historical Kinds and the ‘Special Sciences”. *Philosophical Studies* (95): pp. 45-65.

Mundale (Dissertation)**. “**How Do You Know A Brain Area When You "See" One? A Philosophical Approach to the Problem of Mapping the Brain and its Implications for the Philosophy of Mind and Cognitive Science.” Dissertation.

Mundale (1998)**.** “Brain Mapping,” in William Bechtel and George Graham (eds.), *A Coinpanion to Cognitive Science*. Oxford: Basil Blackwell: pp. 129-139.

Mundale (2002). “Concepts of Localization: Balkanization in the Brain”. *Brain and Mind* (3): pp. 313 – 330.

Nass, Merchant, and Ryan (2008). “*Caenorhabditis elegans* in Parkinson’s Disease drug discovery: addressing an unmet medical need”. *Molecular Interventions* (8): pp. 284 – 293.

Patrissi and Stein (1975). “Temporal factors in recovery of function after brain damage”. *Experimental Neurology* (47): pp. 470 – 480.

Petit, Clark, Ingeholm, and Haxby (1997). “Dissociation of Saccade-Related and Pursuit-Related Activation in Human Frontal Eye Fields as Revealed by fMRI”. *Journal of Neurophysiology* (77): pp. 3386 – 3390.

Place (1956). “Is Consciousness a Brain Process”. *British Journal of Psychology* (47): pp. 44-50.

Polger (2002). “Putnam’s Intuition”. *Philosophical Studies* (109): pp. 143 - 170.

Polger (2004). *Natural Minds*. Cambridge, MA: MIT Press.

Polger (2009). “Evaluating the Evidence for Multiple Realizability”, *Synthese* 167: pp. 457 – 472.

Pons, Garraghty, Ommaya, Kaas, Taub and Mishkin (1991). “Massive Cortical Reorganization after Sensory Deafferentation in Adult Macaques”. *Science* 28 (252), pp. 1857 – 1860.

Putnam (1967). “The Nature of Mental States”, reprinted in Putnam H, *Mind, Language, and Reality*. Cambridge: Cambridge University Press, pp. 429 – 440.

Pylyshyn (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.

Richardson (1979). “Functionalism and Reductionism.” *Philosophy of Science* (46), pp. 533-558.

Sadato, N. (2005). “How the Blind “See” Braille: Lessons From Functional Magnetic Resonance Imaging”, *Neuroscientist* (11), pp. 577 – 582.

Sadato, Pascual-Leone, Grafman, Ibanyez, Deiber, Dold, Hallett (1996). “Activation of the Primary Visual Cortex by Braille Reading in Blind Subjects”, *Nature* (380): pp. 526 – 528.

Sharma, Angelucci, and Sur (2000). “Induction of visual orientation modules in auditory cortex”. *Nature* (404), pp. 841 – 847.

Shapiro (2000). “Multiple Realizations”, *Journal of Philosophy* (97): pp. 635 –654.

Shapiro (2004). *The Mind Incarnate*. Cambridge, MA: MIT Press.

Shapiro (2008). “How to Test for Multiple Realization”, *Philosophy of Science* (75): pp. 514 – 525.

Smart (1960). “Sensations and Brain Processes”. *Philosophical Review* (68): pp. 141-156.

Sober (1999). “The Multiple Realizability Argument Against Reductionism.” *Philosophy of Science* (66), pp. 542-564.

Sur, Garrahgty, and Rowe (1988). “Experimentaly induced visual projections into auditory thalamus and cortex”. *Science* (242), pp. 1437 – 1441.

Ungerleider and Mishkin (1982). "Two Cortical Visual Systems", in D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (eds.), *Analysis of Visual Behavior*. Cambridge, MA: MIT Press, pp. 549-586.

Uttal (2001). *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. Cambridge, MA: MIT Press.

van Essen, Anderson, and Felleman (1992). "Information Processing in the Primate Visual System: an Integrated Systems Perspective". *Science* (255): pp. 419-423.

Van Fraassen (1980). *The Scientific Image*. New York: Oxford University Press.

Von Melchner, Pallas, and Sur (2000). “Visual behavior induced by retinal projections directed to the auditory pathway”. *Nature* (404): pp. 871 - 875.

Wilson, O’Scalaidhe, and Goldman-Rakic (1993). “Dissociation of Object and Spatial Processing Domains in Primate Prefrontal Cortex”. *Science* (260): pp. 1955 – 1958.

Zeki (1977). "Colour Coding in the Superior Temporal Sulcus of the Rhesus Monkey Visual Cortex", *Proceedings of the Royal Society of London,* B197: pp. 195 - 223.