I, Douglas Keaton, hereby submit this original work as part of the requirements for the degree of:

Doctor of Philosophy

in Philosophy

It is entitled:

Realization and Causal Role-Playing: an Essay on the Mind/Body Problem

Student Signature: Douglas Keaton

This work and its defense approved by:

Committee Chair: Thomas Polger, PhD

Carl Gillett, PhD

Robert Skipper, PhD

Christopher Gauker, PhD
Realization and Causal Role-Playing:
An Essay on the Mind/Body Problem

A dissertation submitted to the

Graduate School

of the University of Cincinnati

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in the Department of Philosophy
of the College of Arts and Sciences

by

Douglas Keaton

M.A. University of Cincinnati

May 2010

Committee Chair: Thomas W. Polger, Ph.D.
Abstract:

In this dissertation I argue that one of the central ideas of late twentieth and early twenty-first century philosophy of mind -- that the mind/body relation is the so-called “realization” relation, where realization is explicated as causal role-playing -- is unworkable and ought to be discarded. As a result, a fundamental reevaluation of recent work on the nature of mental states and mental causation is in order. My argument is neither reductionist nor anti-reductionist in intent, since in recent philosophy of mind both reductionists and anti-reductionists have relied upon causal role realization in the construction of their views. Rather, my target is the quasi-behaviorist assumption shared by both reductive and anti-reductive functionalists that the definitive characteristics of mental states are their causal roles; in the memorable phrase of David Lewis, their “syndromes of most typical causes and effects.”

My argument proceeds as follows. In order to identify realization with, or explicate realization as, causal role-playing, it is necessary to identify realized properties such as mental properties with one of the kinds of property that are associated with the causal role-playing relation. I argue that there are four such kinds of property: core realizers, total realizers, and two kinds of role-property that I call A- and B-type role properties. (I show that reductive functionalists attempt to identify realized properties with core or total realizers, while non-reductive functionalists attempt to identify realized properties with one of the two sorts of role property.) But realized properties cannot be identical to any of these four. Realized properties cannot be identical to total realizers or to either type of role property because realized properties are supposed to be causally efficacious but these three kinds of role-related property are not causally efficacious. Core realizers are causally efficacious but realized properties cannot be identical to them because realized properties do not supervene on core realizers.

The argument presented in the previous paragraph constitutes the bulk of my dissertation. As an additional matter I investigate another sort of realization, different from the sort that finds its ancestry in the functionalist philosophies of mind developed by Hilary Putnam or David Lewis. So called “cross-level realization” finds its roots in Jerry Fodor’s use of multiple realization to establish the autonomy of the special sciences and is today employed by many metaphysicians of science. Cross-level realization is not claimed by metaphysicians of science to be explicable as causal role-playing. I argue that philosophers who make use of cross-level realization have not offered any explication of it at all, correct or otherwise, of the sort that causal role-playing purports to be for the sort of realization employed in the philosophy of mind.

I do not propose, in my conclusion, to offer an alternative account of realization. Rather, I argue that the wish for realization is the wish for a metaphysical panacea that makes naturalism metaphysically easy. More technically, it is the wish for a kind of property that strongly supervenes upon lower-order (or lower-level) properties that both do and do not perform requisite legitimating causal duties. This cannot be had; therefore, in particular, the mind-body problem remains unsolved.
Copyright 2010 by Douglas Keaton
Acknowledgements

Asking someone to read your dissertation is a bit like asking them to help you move. The generosity of those who agree is humbling, and leaves you hoping all the more that you didn’t create to much of a mess that they have to help you clean up. In that spirit I would like to apologize to and thank, again, my dissertation committee for their time and their many helpful conversations, criticisms and suggestions. They are Thomas Polger, Christopher Gauker, Robert Skipper, and Carl Gillett. I would like to thank the faculty of the Philosophy Department at the University of Cincinnati for their patience and kindness. I would like to thank my fellow graduate students as well. It is true what they say: most of what you learn in graduate school, you learn from the other grad students.

I would like to thank above all my parents, and, again, my dissertation director, Thomas Polger. Whatever is near-correct in this dissertation got that way as a result of Tom’s guidance and assistance. As I have told anyone who has cared to listen, I could not have found a better director for this dissertation anywhere in the country.

Much of the work for this dissertation was funded by a Taft Dissertation Fellowship, provided by Charles Phelps Taft Research Center at the University of Cincinnati.
# Table of Contents

Part One: Introduction

Chapter One: Introduction 1

Part Two: Same-Level Realization

Chapter Two: Four Theses about Total Realizers 52
Chapter Three: Kim’s Supervenience Argument and the Nature of Total Realizers 73
Chapter Four: Two Kinds of Role Property 103

Part Three: Cross-Level Realization

Chapter 5: Cross-Level Realization 127

Part Four: Conclusion

Chapter 6: Conclusion 180

Bibliography 190
Part One

Chapter One: Introduction

§0

What follows is a dissertation on the mind/body problem. More specifically, it is a dissertation about the most popular current solution to the mind/body problem in analytic metaphysics of mind. According to that solution, the relationship between neurological states of brains and mental states of brains is the realization relation. To put that another way, according to this solution, neurological states of brains realize states of mind. Here, “realize” is a highly technical term that picks out a specific metaphysical necessitation relation; a particular tool in the metaphysician’s tool box; a particular relation next to other relations in the toolbox such as composition, constitution, and identity. Realization is different from these other relations in being much newer, and still, as it were, under development.

In this long introductory chapter I try to accomplish several things. I begin with a discussion of the origins of the idea of realization. As mentioned, realization is a new proposal for a relation in the metaphysician’s toolkit, so the story about origins need not go back very far. In fact, I believe that we need to go back a little further than the standard history of realization would have us go, and that for interesting reasons, but not much further. The discussion of the origins of the idea will take us from the 1950s to the 1960s and it will take up §1-3 of this introductory chapter.

Next, I will take some time to relate a history of further development of the idea of realization from the 1970s to now. At first it was not really clear just what the realization
relation is. It was not really clear just what was being said when it was said that states of the brain “realize” states of mind. But then from the 1970s to now the idea of realization came to be explained in terms of another relation called “the causal role-playing relation.” To say that one property of an object “realizes” another property of the object is to say that former “plays the causal role” of the latter. The story of how this answer developed will take up §4-7 of this introductory chapter. Finally, in §8-10 I will introduce the arguments that constitute the substance of this dissertation.

The primary thesis of this dissertation is: Realization is not and cannot be explicated as the causal role-playing relation. So stated, the thesis looks technical, dry and of interest only to specialists in the metaphysics of mind. However, I hope to make clear in this introduction that the thesis is bold, significant, and of broad import. This, for two primary reasons. First, since the realization relation is the modern-day proposed solution to the mind-body problem, and since causal role-playing is the standard account of just what the realization relation is, the thesis of this dissertation shows that, in fact, nothing coherent is being said when the most common contemporary solution to the mind/body problem is put forward. This means in turn that a fundamental reevaluation of the results of the past forty years in the metaphysics of mind, is in order. Second, the idea of realization has, in virtue of its “success” in the philosophy of mind, come to be used in the metaphysics of science and indeed in the most influential current forms of general physicalist metaphysics. That is, the current ways of making sense of the way the human world can exist in the physical world described by the sciences make crucial use of the idea of realization. The thesis of this dissertation again shows that much fundamental reevaluation is in order. This, I think, is good: it clears the way for new work to begin in a new century.
Sydney Shoemaker’s 2007 book *Physical Realization* was the subject of an “author meets critics” session of the 2009 Pacific Division meeting of the American Philosophical Association. Louise Antony, Jaegwon Kim, and Andrew Melnyk, as the “critics,” presented papers. In one session were gathered together four of the most prominent current researchers into the nature and use of the so-called “realization relation.”¹

I would like to begin this dissertation by discussing some remarks from Shoemaker’s book and Kim’s paper, from that session, on the origins of the notion of realization. Though Kim had for his part told a similar story before, in the first chapter of his *Mind in a Physical World*, this newer version is of special interest because of the occasion. Here were two elder statesman in the philosophy of mind discussing the origins of what is arguably currently the most important idea in the field, the idea that the mind/body relation is the realization relation. In the remainder of this section I discuss Shoemaker’s take on the history of realization as presented in *Physical Realization*, and then in §2 discuss Kim’s addenda.

Shoemaker observes that “the notion of realization figures prominently in recent discussions of physicalism.” Previously the notion of realization had been around, and important, but it did not figure prominently in discussions of physicalism. Realization, Shoemaker tells us, was originally developed and deployed to address the mind/body problem in particular, not physicalism in general. Very specifically, realization was originally developed to combat the Identity Theory -- the theory developed in the 1950s in the philosophy of mind that

¹ The papers were later published in the Philosophical Quarterly. See Antony (2010), Kim (2010), Melnyk (2010), Shoemaker (2010).
mental states are strictly identical (type identical) to physical states. The range of realization’s uses expanded later, as philosophers came to understand that the reasons to study the notion of realization went beyond the philosophy of mind. “It is arguable,” Shoemaker writes in expression of that broader understanding, that the notion of realization “provides the most revealing characterization of physicalism itself: physicalism, we can say, is the view that all states and properties of things, of whatever kind, are physical or physically realized” (Shoemaker, 2007, p. 1).

When the story of realization is told in this way -- and this is the standard way to tell it, with the notion of realization beginning in the philosophy of mind and expanding outward -- it can seem an odd tale. It can seem as though what happened is that a lot of philosophers saw a proposal for the mind/body relation and decided that not just minds but everything can be explained in that same way. Thus, the idea that Shoemaker is fronting, when he says that physicalism may be fruitfully understood as the thesis that all properties and states are either physical or physically realized, would seem to be re-statatable as follows: physicalism may be fruitfully understood as the thesis that all states and properties of things, of whatever kind, are either physical or else they stand in the same relation to physical properties and states that mental states stand in to brain states. This is the picture of physicalism that we get from the standard history that Shoemaker is relating. Shoemaker seems to be saying that no matter what stripe of physicalism one adopts, one is adopting a position that may be fruitfully understood as the thesis that all properties and states stand in the same relation to physical properties and states that mental states stand in to brain states. That seems like a bold thing to say.

---

2 See Smart (1959); and for a look back at Smart’s article and an overview of the current status of the Identity Theory, Polger (forthcoming).
This characterization of physicalism would be rendered less bold and more plausible if it turned out that “realization” named a class of relations rather than one very specific relation, or anyway if “realization” turned out for whatever reason to be an innocuously non-specific term. And indeed for a moment Shoemaker seems to adopt that view. Shoemaker notes that “one of the dictionary definitions of the verb ‘realize’ is ‘make real’” (p. 1). If that is all that is meant by “realize” then it might indeed be that physicalism can be non-committally but interestingly characterized as the view that all properties and states are either physical or realized by -- made real by -- physical properties and states just as mental states are made real by brain states. To commit oneself to a physicalism grounded in realization in this sense need not be to commit oneself to anything more than a general ontological naturalism.3

But that turns out not to be what Shoemaker has in mind. The “make real” sense of “realize” is not the sense that is in philosophical use; it is not the sense that illuminates physicalism:

One of the dictionary definitions of the verb “realize” is “make real.” But the ordinary notion that the dictionary definition captures has to do with the fulfillment of plans, intentions, desires, etc.; as we might put it, what realizes a desire is what makes real the intentional object of the desire. And this is not the notion that is in play when a philosopher speaks of mental states as realized in physical states. (Shoemaker, 2007, p. 1)

Shoemaker compares the “dictionary sense” of realization in which a desire to see the Taj Mahal is made real by actually seeing the Taj Mahal, to the “philosopher’s sense” of “realization” in which the desire to see the Taj Mahal is realized in a brain state. Shoemaker

3 Ontological naturalism and this kind of physicalism are not quite coextensive -- since one could hold that it could turn out that the natural sciences show us that there are naturally occurring properties that are not helpfully thought of as being made real by physical properties -- but the two views are nonetheless pretty close. Kissing cousins at least.
goes on to say that, in contrast to the dictionary sense, “realization,” in the “philosopher’s sense,”
is a “term of art” that does not have much directly in common with ordinary usage. “Still,”
Shoemaker adds, “defining it as ‘make real’ is a good first stab at capturing its meaning” (p. 2).

This sort of move tends to be a key moment in discussions of the notion of realization.
Shoemaker seems to be doing his level best both to ground the philosopher’s notion of
realization in an ordinary sense of “make real” but also to keep the philosopher’s notion free
from that ordinary sense. This creates the following tension: if “realization” in the philosopher’s
sense is just a “term of art” then it is strictly irrelevant that the term is homonymous with an
ordinary-language term that means “make real.” As Kim will later say, because “realization” is a
“term of art” then we are free to judge construals of it or explications of it strictly by their
“philosophical usefulness” with no care as to how well a given construal or explication captures
or keeps in touch with pre-philosophical notions of what it is for one thing to make real another
thing (Kim, 2010, p. 103).

On the other hand, I add, there is a good reason for the philosopher to want to keep
philosophical terms that begin in ordinary usage in contact with ordinary usage. The good
reason is that this contact can underwrite expansions of the use of the concept into new domains.
Part of my point here is that if Shoemaker is right that realization began in the philosophy of

---

4 Two points. First, Polger distinguishes six pre-philosophical senses of “realize” (Polger, 2007, p. 234). One of
them strikes me as being right at the cusp of ordinary use and the philosophical term of art: the sense of “realize” in
the sentence “Rodin’s The Thinker was realized in wax and bronze” (ibid.). It is this sort of sense that I will appeal
to later on.
Second, I think there is a small confusion in the Taj Mahal example. I think ordinary usage would not have us say
that the desire to see the Taj Mahal can be realized by taking a trip to Agra, India and actually seeing the Taj Mahal.
I think ordinary usage has it that what is realized in that scenario is not the desire to see but the desired seeing. But
if this is so, then even Shoemaker’s example provides a perfectly ordinary sense of “realize” that is much closer to
the philosopher’s sense of “make real” than Shoemaker allows. The seeing in this scenario is no longer aspirational
but actual, a seeing that the trip and the act of looking have made real. This will be of some import later in this
introduction when I argue that one may abandon any particular explication of the philosopher’s notion of realization
while still keeping a sense of realization that is only sort-of-technical, one that has contact with ordinary usage while
still having in hand a notion of realization that is of use in characterizing physicalism or naturalism.
mind and expanded out to general metaphysics, the expansion has at the very least an air of arbitrariness and might be flat inexplicable, if realization is merely a term of art. My point will become clearer as Shoemaker continues:

The brief history of the notion of realization is entangled with the history of functionalism in the philosophy of mind. The idea that mental states can be multiply realized figured centrally in Hilary Putnam’s rejection of type physicalism in his seminal paper “The Nature of Mental States” [(Putnam, 1975b)]. It also figures prominently in Jerry Fodor’s “Special Sciences” (Fodor, 1974). I do not know who was the first to use the word “realizer” for what does the realizing. (Shoemaker, 2007, p. 2)

In “The Nature of Mental States” Putnam spoke of the physical realization of probabilistic automata. He also spoke of the physical realization of sense and motor organs -- a point often overlooked.\(^5\) If we read Putnam as making use of a term that may apply to sense organs as well as to mental states then it would seem not to be a term of art but a term in contact with an ordinary and quite general sense of “make real.” But if we read Putnam as introducing “realization” as a term that applies only to Probabilistic Automata then presumably it is strictly a “term of art.” I believe that Shoemaker has in mind the latter reading, though that leaves it

---

\(^5\) Putnam wrote:

I shall assume the notion of a Probabilistic Automaton has been generalized to allow for ‘sensory inputs’ and ‘motor outputs’ -- that is, the Machine Table specifies, for every possible combination of a ‘state’ and a complete set of ‘sensory inputs’, an ‘instruction’ which determines the probability of the next ‘state’, and also the probabilities of the ‘motor outputs’ . . . I shall also assume that the physical realization of the sense organs responsible for the various inputs, and of the motor organs, is specified, but that the ‘states’; and the ‘inputs’ themselves are, as usual, specified only ‘implicitly’ -- i.e. by the set of transition probabilities given by the Machine Table.

Since an empirically given system can simultaneously be a ‘physical realization’ of many different Probabilistic Automata, I introduce the notion of a Description of a system . . . (Putnam, 1975b, p. 434)

The fact that Putnam was willing to speak of the ‘physical realization’ of sense organs and motor organs, in addition to the more familiar ‘implicitly’ specified realization of the machine table, leaves it much less clear just what sort of term of art Putnam took himself to be introducing, if any. (Note too that the rhetorical implication is that even though the “physical realization” of the Machine Table is “specified only implicitly,” the physical realization of sense and motor organs may be made explicit by the Description. This indicates that “realization” for Putnam does not imply implicit specification. This, again, renders dubious the claim that Putnam took himself to be introducing “realization” as a highly technical and specific “term of art.”
unexplained just why a highly technical term that applies to Machine Table states ought to find application in discussions of the special sciences generally.

Another point: we just saw Shoemaker speak of Putnam’s introduction of the notion of realization into the philosophy of mind and then, in the same breath, as it were, of Putnam’s introduction of the idea of “multiple realization” of mental states. This is correct as far as it goes (Putnam did introduce both notions) but Shoemaker’s wording presents a good opportunity for me to make a key distinction that he is eliding. For the purposes of introducing this dissertation I should draw a bright, clear line between, on the one hand, attempts to understand the philosophical notion of “realization” and, on the other hand, attempts to understand the philosophical notion of “multiple realization.” It was specifically the idea of multiple realization that Putnam employed in his rejection of type physicalism. Realization is a different idea, one that predates multiple realization. Putnam proposed the realization relation as a replacement for the Identity Theory once the Identity Theory was rejected with the help of the notion of multiple realization. I will now spend a few paragraphs on the distinction between realization and multiple realization and then return to Shoemaker’s discussion of the origin of the notion of realization.

Here is the difference between researching the realization relation and researching multiple realization. If we engage in the project of trying to understand the realization relation we are asking after the nature of a metaphysical posit -- a relation that, it is supposed, may obtain between properties; it is offered as an explanation of the presence in the world of certain sorts of properties whose presence would otherwise remain (metaphysically) mysterious. We are trying to understand just what sort of explanation is being offered for the presence of the otherwise
mysterious properties and we are trying to understand whether the explanation is a good one. On the other hand, if we engage in the project of trying to understand multiple realization we are asking after the apparent fact that entities that seem to be of the same kind when looked at from the point of view of one science may seem to be of clearly different kinds when looked at from the point of other, particularly more basic, sciences.

If we want to know how and whether it could be the case that the kinds of things that are studied in one science could be regarded as very different kinds of things by another science, and under what circumstances, then what we want to know about it is possibility and scope of multiple realization. If, on the other hand, we want to know what it is that explains this multiple realization, the answer might be “realization.” More specifically, the answer might be some particular version of the philosopher’s quasi-technical notion of realization. Or, the answer might not be anything recognizable as the philosopher’s quasi-technical notion at all. Whether multiple realization occurs does not stand or fall with the success of any one attempt to understand what the realization relation might be. Indeed, whether multiple realization occurs might not stand or fall with the success or failure of all attempts to understand any version of the philosopher’s notion of realization.

Here is another way to make the point. Thomas Polger has pointed out that Putnam’s multiple realizability argument is supposed to be a definitive critique of the Identity Theory in the philosophy of mind. It is also supposed to be evidence that the mind-body relation is the realization relation. But whether the multiple realizability argument succeeds in defeating the Identity Theory is an issue that is separable from the issue of whether the multiple realizability argument provides sufficient reason to believe in realization. Philosophers such as Putnam may
use multiple realization as *defeasible evidence for* the existence of (some specific version of) the realization relation, but that already shows that multiple realization and the (or “the”) realization relation are not the same thing.⁶

Again, I stress this distinction because the current dissertation is about *the realization relation*, not *multiple realization*. I want to examine explications of the realization relation -- specific ways of understanding the metaphysical posit that does so much heavy lifting in current philosophy of mind and metaphysics of science. I am not, for current purposes, interested in the ins and outs of *multiple realization*, of the suggestion that there is, to wit, more than one way to build a mind, or a glacier.

Here is where we are. Shoemaker writes that the notion of realization was introduced into the philosophy of mind by Hilary Putnam in 1967 as a way of understanding the relation between mental states and brain states. The term “realization” has an ordinary sense but the term as Putnam introduced it was a term of art having little to do with that ordinary sense. I have argued that this line of interpretation leaves it unclear just what underwrites the expansion of the use of realization from the philosophy of mind to more general metaphysics -- specifically to the metaphysics of science, as in Fodor’s 1974 “Special Sciences” paper and after. Indeed, Shoemaker’s line of interpretation leaves it unclear just what underwrites passages such as the following, which Shoemaker wrote directly after the previous block quote:

> While discussions of realization have nearly always been discussions of the realization of mental properties, it is important to remember that the application is much broader than this. As noted earlier, one can speak of colors as realized in spectral reflectances. One can speak of the chemical and physical

---

⁶ If indeed it is legitimate to use multiple realization as evidence *at all* for the existence of the realization relation. See Polger (2007, 2008).
realization of biological properties, and of the mechanical, hydraulic, electronic, etc. realization of properties like being a braking system, being a clock, etc. (Shoemaker, 2007, p. 3)

Now we turn to Kim’s friendly addenda to Shoemaker’s remarks on the history of realization. Kim finds an earlier occurrence of realization, in a previous paper by Putnam. This will not significantly affect the exegetical or historical/interpretational problems I have raised in this section. Kim also suggests that it is remarkable that the notion of realization went for as long as it did (until the late 1990s) before philosophers started asking in a philosophically serious way just what the realization relation is.

§2

In his paper for the 2009 APA Pacific symposium on Shoemaker’s book, Kim writes:

Perhaps I can add a bit more light on the history of realization. I believe the concept, and the term, were first introduced by Putnam, in an earlier paper, “Minds and Machines” (Putnam 1960) [(Putnam, 1975a)], in which we find the following passage:

“In particular the ‘logical description’ of a Turning Machine does not include any specification of the physical nature of those ‘states’ [a Turing machine’s ‘internal states’ [Kim’s brackets]] . . . In other words, a given ‘Turing Machine’ is an abstract machines [sic] which may be physically realized in an almost infinite number of different ways.” (Putnam, 1975a, p. 371) [Kim’s italics]

This is the first occurrence of the idea of physical realization that I know of. Apparently, Putnam himself thought so, too, as he writes, later in the same paper (Putnam, 1975a, p. 373):

“The functional organization (problem solving, thinking) of the human being or machine can be described in terms of the sequences of mental or
logical states respectively . . . without reference to
the nature of the ‘physical realization’ of these
states.

It is plausible to suppose that Putnam put quotes around “physical
realization” with the thought that here he was introducing a
technical neologism. (Kim, 2010, pp. 101-102)

Kim is here suggesting that the concept of realization -- clearly referring to the
realization relation as opposed to multiple realization -- and the term “realization” were
introduced in 1960 by Hilary Putnam, in a paper that draws an analogy between the nature of
computational states in machines and mental states in persons.\(^7\) Kim’s evidence that Putnam’s
use is originary (or at least that Putnam takes his use to be originary) is that Putnam put “physical
realization” in quote marks. But then Kim adds, “It is somewhat remarkable, however, that he
offers no explanation of what he means by this new term, apparently counting on the reader’s
intuitive understanding” (p. 102).

But this is unsatisfactory for two reasons. First, the fact that Putnam provided no
explanation for the use of the term “realization” seems to be good evidence that Putnam did not
take his use of “realization” to be originary; that he did not take himself to be “introducing a
technical neologism.” The lack of an explanation for the term is better evidence for the
interpretation that Putnam did not take himself to be introducing a neologism than is the
presence, noted by Kim, of scare quotes, for the interpretation that Putnam was introducing a
neologism. Second, at the time we are discussing, Putnam’s writing style involved putting
quotes around terms frequently. Consider the long quote from “The Nature of Mental States” in
footnote 5. In one paragraph Putnam puts quotes around not just “physical realization” but also

\(^7\) An analogy, not an identity. Putnam would not claim that the mind/body relation was literally the same thing as the
computational implementation relation until “The Nature of Mental States.” See Piccinini (2004) for a good, brief
history of functionalism from the point of view of computationalism.
“sensory inputs,” “motor outputs,” “instruction,” “state,” and “implicitly.” As far as I know, no one interprets Putnam as introducing all of these terms as technical neologisms.

Let us take a step back for a moment. Recall that we are discussing the origins of one of the most important terms and concepts in twentieth and (so far, at any rate) twenty-first century philosophy. Kim and Shoemaker seem to be defending the idea that the concept is sui generis, introduced in a fit of genius by Hilary Putnam in the 1960s. That is certainly possible, but Kim is right that if that is right then it is nearly shocking that Putnam did not pause or slow down to let us know what he was talking about, when he said that logical and mental states, and later on that sense and motor organs, have “physical realizations.” It is also amazing, under this interpretation, that so many philosophers for such a long time took themselves to know what Putnam was talking about. Kim, continuing directly:

It is even more remarkable that no one seems to have raised any questions about how “realization” was to be understood—all this while the term was gaining quick currency through the rest of the century, figuring in some of the most important claims and arguments in philosophy of mind, metaphysics, and philosophy of science in the “postpositivist” era. Without the idea of “realization”, or “multiple realization”, the face of philosophy of mind and of science today would look very different indeed. (Kim, 2010, p. 102)

Yes. But, again, this is a reason to believe that Putnam was not merely “introducing a technical neologism” that had no previous currency in philosophy. Kim adds in a footnote that recently a younger generation of philosophers has begun to ask just what realization is. But that does nothing to allay the worry that the previous generation could not have been so thick as to let a concept central to so much work in late twentieth century philosophy pass by unexamined. In

---

8 Kim writes, “That is, until quite recently. See works by Lawrence Shapiro, John Bickle, Ronald Endicott, Carl Gillett, Lenny Clapp, Tom Polger, and others” (Kim, 2010, p. 102, f. 4).
particular, it leaves central and crucial transition from talk of the realization of machine tables to talk of the realization of properties completely unexplained:

As Shoemaker says, the dictionary meaning of “realize” is something like “make real”. I believe this fits Putnam’s original usage quite well: a physical computing machine “makes real” or “brings into concrete reality” an abstractly characterized Turing machine, a mathematical entity. In this sense, what is realized are things like plans, designs, blueprints, and the like, and realization is a relation between concrete objects in the world and abstract items like Turing machines, plans, and designs. As we all know, this has changed: we now think of realization primarily as a relation between properties; we talk about neural/physical properties realizing mental properties, physicochemical properties realizing biological properties, and microphysical properties realizing observable macroproperties of physical objects. (Kim, 2010, p. 103)

That is key. When philosophers nowadays talk about the realization relation they are speaking of properties, of characteristics possessed by material objects. That is, both of the relata are such characteristics. On the Putnamian understanding a characteristic or property may realize a machine table state, but the state itself is not a characteristic or property; it is an abstract object, a machine table. On the contemporary understanding, the thing that is “made real” is a characteristic of a material object just as much as is the thing that does the real-making. To what extent the realized characteristic, such as “pain” or “the property of being in pain,” to use the standard example in contemporary parlance of a realized mental property, is itself a material characteristic of the material object that has it, is one of the issues debated by philosophers who study realization. But that realized properties are properties of entities found in the natural world, is common coin.

What would be helpful would be to find an early discussion of realization that (i) clearly tied realization in philosophy to the colloquial sense of “make real” and (ii) clearly tied talk of
realization to concerns about general philosophical naturalism. It would be helpful if this discussion predated Putnam’s more specific use of “realization” in the then-new functionalist philosophy of mind. If found, such a discussion would help explain everything here that seems mysterious. It would explain why Putnam did not feel the need to explain his use of “realization”; it would explain why talk of realization was allowed to spread without comment from the philosophy of mind the more general metaphysics; it would explain why talk of realization could go from talk of machine tables to talk of properties found in nature. It would explain all this because it would show that realization really did not begin with talk of machine tables, but with talk of naturalism. It’s just that nowadays no one remembers it.

In the next section I present such a discussion.

§3

Ernest Nagel’s 1954 Presidential Address to the Eastern Division of the American Philosophical Association was devoted to the defense of philosophical naturalism against the charge that it is small-minded, scientistic, methodologically provincial -- at any rate, that naturalism constitutes a surrender of the traditional philosophical goal of understanding the world as a whole. Naturalists are not system builders, the charge went:

The past quarter century has been for philosophy in many parts of the world a period of acute self-questioning, engendered in no small measure by developments in scientific and logical thought, and in part no doubt by fundamental changes in the social order. In any event, there has come about a general loss of confidence in the competence of philosophy to provide by way of a distinctive intellectual method a basic ground-plan of the cosmos, or for that matter to contribute to knowledge of any primary subject-matter except by becoming a specialized positive science and subjecting itself to the discipline of empirical inquiry. (Nagel, 1954, p. 6)
The worry was that naturalism as philosophical method or outlook or attitude constituted a retreat from philosophy. Nagel’s rebuttal was that naturalism was two-fold. First, naturalism was for Nagel, whatever else it might have been, a species of philosophical honesty.

With Santayana, I prefer not to accept in philosophic debate what I do not believe when I am not arguing; and naturalism as I construe it merely formulates what centuries of experience have repeatedly confirmed. At any rate, naturalism seems to me a sound generalized account of the world encountered in practice and in critical reflection, and a just perspective on the human scene. (Nagel, 1954, p. 7)

Second, despite claims that naturalism is merely a method, empty of philosophical content, and a parasitic one at that, naturalism does make positive claims. It is here that Nagel begins to bring in the notion of realization. As we will see, after explicitly using the term “realization” in a discussion of the theses of naturalism, Nagel will go on to make claims about the naturalistic view of the world that are strikingly prescient of the sorts of claims that friends of realization in the metaphysics of science have been concerned to make in more recent years.

Nagel writes that there are two theses of naturalism. Both are of interest here. The first stresses the continuity of nature:

Two theses seem to me central to naturalism as I conceive it. The first is the existential and causal primacy of organized matter in the executive order of nature. This is the assumption that the occurrence of events, qualities and processes, and the characteristic behaviors of various individuals, are contingent on the organization of spatio-temporally located bodies, whose internal structures and external relations determine and limit the appearance and disappearance of everything that happens.

The second stresses the discontinuity of nature:

The second major contention of naturalism is that the manifest plurality and variety of things, of their qualities and their functions,
are an irreducible feature of the cosmos, not a deceptive appearance cloaking some more homogeneous “ultimate reality” or transtemporal substance, and that the sequential orders in which events occur or the manifold relations of dependence in which things exist are contingent connections, not the embodiments of a fixed and unified pattern of necessary links. The existential primacy of organized matter does not make illusory either the relatively permanent or the comparatively transient characters and forms which special configurations of bodies may possess. (Nagel, 1954, p. 8)

With the second thesis, Nagel seems to have been concerned to combat something like the doctrine of internal relations, as well as an overweening reductionism. (The concern with the doctrine of internal relations is dated. The concern to combat brash reductionism remains current.)

But how to connect these two theses, these two apparently antagonistic commitments to the continuity and discontinuity of the natural world? Enter talk of realization, in a passage that is both quite lovely and that ties together many of the consequences of the two theses of naturalism.

Naturalism does not maintain that only what is material exists, since many things noted in experience, for example, modes of action, relations of meaning, dreams, joys, plans, aspirations, are not as such material bodies or organizations of material bodies. What naturalism does assert as a truth about nature is that though forms of behavior or functions of material systems are indefeasibly parts of nature, forms and functions are not themselves agents in their own realization or in the realization of anything else. In the conception of nature's processes which naturalism affirms, there is no place for the operation of disembodied forces, no place for an immaterial spirit directing the course of events, no place for the survival of personality after the corruption of the body which exhibits it. (Nagel, 1954, pp. 8-9; italics original; my boldface)

Let us tread carefully here. It appears that Nagel is making the following claims:
(N1) It is not the case that all existing things are material things.
(N2) Dreams and joys, plans and aspirations have been noted in experience.
(N3) Dreams and joys, plans and aspirations are not material bodies.
(N4) Dreams and joys, plans and aspirations are not organizations of material bodies.
(N5) Forms of behavior and functions of material systems are parts of nature.
(N6) Forms and functions are not responsible for their own realization.
(N7) Forms and functions are not responsible for the realization of anything other than themselves.

I think that we are intended to draw the following (non-deductive) conclusion from (N1) and (N2):

(N2a) Dreams and joys, plans and aspirations, exist.

Further, it seems to me that the transition from (N4) to (N5) is inexplicable unless the evident existence of non-material things such as the ones mentioned in (N2a) are somehow connected to the “forms of behavior and functions of material systems” mentioned in (N5). I propose that for current purposes -- that is, to help us to understand Nagel’s intent -- we simply identify them. It seems to me that Nagel’s suggestion is that non-material things of the sort mentioned in (N2a) just are “forms of behavior and functions of material systems.” Or at any rate this is close enough to what he appears to have in mind to go on with.

(N4a) Dreams and joys, plans and aspirations are forms of behavior and functions of material systems.

(N4a) renders (N5) relevant; together, (N4a) and (N5) resolve the tension between (N2a) and the conjunction of (N3) and (N4).

From here, it seems pretty evident how to schematize the remainder of the line of thought that Nagel is pursuing.
Forms and functions of material systems are realized by what is material, viz. material bodies and organizations of material bodies.

From (N4a) and (N5a) it follows that:

Dreams and joys, plans and aspirations are realized by what is material, viz. material bodies and organizations of material bodies.

But from (N5b) and the conjunction of (N6) and (N7) a troubling consequence follows:

Dreams and joys, plans and aspirations are not responsible for their own realization or for the realization of anything else.

Since dreams and joys, aspirations and desires are not material bodies, they are not part of Nagel what called “the executive order of nature.” They exist because they are forms and functions of systems that are parts of the executive order of nature. But as such, dreams and joys, aspirations and desires do not make anything happen in the world, they do not affect the executive order nature from the top down, as it were. Here, rather remarkably, we have the problem of causal exclusion twenty years before Norman Malcolm would write “The Conceivability of Mechanism” (Malcolm, 1968) and forty years before the decade in which Kim would make the exclusion problem central to the study of functionalism in the philosophy of mind (Kim, 1993).

But in a more general sense, here we have the notion of “realization” used to tie together two theses of what Nagel refers to as two positive theses of “naturalism” -- what today we might call two theses of non-reductive physicalism: the thesis that the natural world is continuous inasmuch as everything that happens in the natural world is due to events occurring in the executive order of things, and the thesis that the natural world in not continuous inasmuch as
“the manifest plurality of things, of their qualities and their functions, are an irreducible feature of the cosmos.”

Additionally, though it might appear that Nagel means to connect talk of realization to mental states alone, as Putnam would appear to do a few years later, I think that Nagel’s use, as examples, of “modes of action, relations of meaning, dreams, joys, plans, aspirations,” are just that: examples. In other parts of the address to the APA it is clear that he means to include other sorts of things than mental states in the irreducible plurality of existents that make up the natural world.

The existential primacy of organized matter does not make illusory either the comparatively transient characters and forms which special configurations of bodies may possess. In particular, although the continued existence of the human scene is precarious and is dependent on a balance of forces that doubtless will not endure indefinitely, and even though its distinctive traits are not pervasive throughout space, it is nonetheless as much a part of the “ultimate” furniture of the world, and is as genuine a sample of what “really” exists, as are atoms and stars. There undoubtedly occur integrated systems of bodies, such as biological organisms, which have the capacity because of their material organization to maintain themselves and direction of their characteristic activities. (Nagel, 1954, p. 9)

Here Nagel seems to include biological organisms in the irreducible totality of things, and though it is a bit of a stretch, I think it is not a great stretch, to suggest that Nagel had in mind biological entities -- as well as entities described in other special sciences -- in the earlier passage as well, when he spoke of realization in connection with the “variety of things, of their qualities and their functions” that “are an irreducible feature of the cosmos.”

So what we have here is an example of the use of “realization” that predates Putnam’s use of “realization” in “Minds and Machines.” It occurs in a prominent publication and the use of
the term “realization” is at least as highly featured in this publication as was Putnam’s use of “realization” in “Minds and Machines.” It is evident that Nagel intends “realization” in the sense of “make real” and he is using it in a philosophical context. The address in which the use occurs is, crucially, a defense of a general naturalistic picture of the world; the address is not particularly devoted to the philosophy of mind, though Nagel’s most prominent examples of realized entities are what we would now call mental states or processes. “Realization” is used to defend a view that Nagel calls “naturalism” but which contains positive elements that appear strikingly like a version of non-reductive physicalism that we might see defended by non-reductive physicalists today.

Thus, we have a candidate for an originary use of “realization” that predates Putnam and occurs in a broader context than just the philosophy of mind. This explains both why Putnam felt he could use term “realization” in a narrower context without explanation in his papers in philosophy of mind in the 1960s and it explains why talk of realization spread out to other areas of philosophy beyond the philosophy of mind in the years after Putnam’s publications with such ease.

Nagel’s 1954 Presidential Address also helps to answer what for me is a key question: Do we need to throw out the concept of realization if it can be shown that the way that the concept was developed in the philosophy of mind from the 1960s to today was fundamentally misguided? If “realization” is merely a “term of art” or a “technical neologism” developed specifically to aid in the construction of specific views in functionalist philosophy of mind then presumably the concept of “realization” will lose its use when or if the views it was developed to aid are discarded. On the other hand, if “realization” has a quasi-colloquial sense that is tied in a very
general way to the defense of philosophical naturalism -- a view that we might, with some equivocation, depending upon the context, call “non-reductive physicalism” -- then presumably the notion of “realization” will be able to withstand the dismantling of any sub-specialty’s parochial explication of it.

Which is just as well, since what I intend to do in this dissertation is to establish that one particular technical explication of the notion of “realization” -- by far the most common one, since Putnam and Lewis began to employ the term in the philosophy of mind -- is fundamentally misguided and ought to be discarded.

I do not think that Nagel’s defense of naturalism needs to be discarded, however, and I am glad to discover that I can attack a technical explication of the realization relation without attacking the root notion of “realization” as Nagel seems to have understood it.\(^9\) My target is not naturalism but functionalism, in particular causal role functionalism. One way to put my project, then, is this: I wish to rescue aspects of Nagel’s 1954 view of the world with its irreducible plurality of things from functionalism. I wish to free non-reductive physicalism from the notion that it must be synonymous with causal role functionalism.

§4

By far the most common view currently held in analytic metaphysics of mind about the nature of mental states is causal role functionalism. According to it, states of mind such as

---

\(^9\) I say in the text that “I do not think that Nagel’s defense of naturalism needs to be discarded.” By the time we get to the conclusion we will see that, in fact, I do think that Nagel’s defense of naturalism needs to be discarded. But for now it will help if I present the philosophical dialectic as follows: causal-role functionalism developed as an attempt to work-out the implications of Nagel’s view. It is possible for causal role functionalism to fail even if what it attempted to explicate remains. So to that extent my dissertation is not an attack on Nagel’s view. In the conclusion, however, I will say something about the philosophical roots of the mistakes that I think I am uncovering, and for that I will have to criticize Nagel’s defense of naturalism.
beliefs, itches, hopes, seeing the wet wheelbarrow glazed with rainwater beside the white chickens, and hearing a Beethoven movement, have unique and definitive causal roles. For a person to be in a state of mind such as one of the ones just listed is for that person’s brain to be in a neurological state that plays the causal role that is definitive of that state of mind.

A quick janitorial matter before we proceed: it is currently typical in this field to speak of properties rather than states. One says that neurological properties play the role of mental properties (in humans, at least). This can get rather circuitous when philosophers adopt the view that it is not properties as such that play the causal role of mental states in persons on particular occasions, but rather instances of those properties. This can be circuitous because one definition of property instance is state of affairs, where a state of affairs is an ordered triple consisting of an object, a property, and a time. If one thinks that -- in humans at least -- it is neurobiological property instances that play mental causal roles then one evidently might as well have just stuck with the term “states” in the first place. In any event, I shall follow the usual policy of referring to “properties” unless greater precision is needed.

David Lewis is arguably the founder of causal role functionalism in the philosophy of mind. According to an early paper by Lewis, every experience is definitively characterized by its causal role, what he called its “syndrome of most typical causes and effects” (Lewis, 1983a, p. 100). The standard example to give of such a definition-by-syndrome is a truncated version of a

---

10 This is how Sydney Shoemaker defines “property instance,” at any rate, and his metaphysics of properties (he calls it “the causal theory of properties”) is used by philosophers interested in realization as much as any. See Shoemaker (2003) for the causal theory of properties and Shoemaker (Shoemaker, 2007, p. 21, f. 3) for, specifically, Shoemaker’s definition of “property instance.”

11 The reason it is just “arguable” rather than “obviously true” that Lewis is the founder of causal role functionalism is that Lewis developed and defended a reductive version of causal role functionalism while most later causal role functionalists were (and are) non-reductivists. If anyone is to get credit for showing that causal role functionalism could be given a non-reductive variant, and therefore if anyone deserves credit for being a co-founder of causal role functionalism, it is Block, due to Block (1980).
definition of “pain” in terms of what we imagine is the syndrome of pain’s most typical causes and effects.

\[
\text{Pain} =_{df} \text{the mental state that is caused by inter alia skin damage and that causes inter alia wincing.}^{12}
\]

It might be wondered why anyone would think that the definitive characteristic of pain is its syndrome of most typical causes and effects. Surely, one might want to say, it is the feeling of pain that is definitive of pain, not what pain does or what causes it. To such an objection I have no philosophical reply; my goal here is not to defend causal functionalism but to describe it and, indeed, eventually to refute it. But I believe that the right historical reply, anyway, begins by asking the questioner to recall that causal role functionalism was developed shortly after the demise of behaviorism. The philosophers who developed functionalism thought that there were elements of behaviorism worth saving. Thus Lewis described his view as an application of Sellar’s myth of our Rylean ancestors, a myth that explains “the plausibility of behaviorism” (Lewis, 1999a, p. 259).

Staying with the connection between functionalism and behaviorism for a moment, a standard functionalist critique of behaviorism is that behaviorism must rely upon simple input-output algorithms to get from stimulus to response; behaviorism cannot appeal to more complex causal intermediaries in between stimulus and response because that would be going inside the head -- a methodological move which to a behaviorist is cheating. But this means that a stimulus-response definition of a mental state may never appeal to other mental states. Causal role functionalists believe that mental states can only be defined causally if some of their causes

\[^{12}\text{Sometimes causal role functionalists will define “being in pain” rather than “pain.” I think there is no special reason for this, other than that they consider it easier, in English, to define “being in pain” via a causal role formulation than to define “pain.”}\]
and effects are allowed to be other mental states. Thus, to use an example from Jaegwon Kim, for a ham sandwich to cause you to desire to eat it, you must believe it to be a ham sandwich (Kim, 2006, p. 123). The belief figures essentially in the causal characterization of the desire. Once this line of thought is played out we get the result that causal role functionalism must define all mental states at once if it is going to define any of them. This in fact is what Lewis proposes.

Causal role functionalism in the philosophy of mind relies upon the possibility of providing a complete causal theory of psychology -- a theory that will predict, account for, and define every mental state of any entity with a mind like a human mind. This (merely aspirational, of course) theory is typically called in the literature “the theory T,” with the “T” boldfaced -- a tradition begun in Lewis (1999a). The idea is to treat every mental state term that occurs in T as a theoretical term, a term that gets its meaning entirely from content it derives from T itself. T defines the mental terms all at once; the definitions are interdependent. To do this T employs what Lewis called “O-terms” for “old terms” -- terms that we understood before the introduction of the theory T. It is the O-terms (and the logical terms, if they are not counted as o-terms) that provide the empirical/logical content of the theory and the content of the definitions that are derivable from the theory. It is simplest to think of the O-terms as specifying the stimuli and responses that must be the most typical ultimate causes and effects, however proximal or distal, of any mental state.\textsuperscript{13} Again, since T provides the causal specification of every mental state at once all of the defining causal roles are interwoven. Indeed, any given

\textsuperscript{13} That is to say, a given mental state -- say, ennui -- might be very interior, without having much directly to do with external causes and without having much to do with behavioral effects. Nevertheless, no matter how much mental pinballing around in the skull is needed there must be an eventual suite of external causes (which is to say referents of o-terms in T that are partial distal causes of, say, ennui, even if they are not external to the body) and, similarly, a suite of behavioral effects.
definition will be, in effect, a rewrite of the entire theory with the definiendum moved “out front.”\textsuperscript{14}

So far we have seen something of how causal role functionalism characterizes mental properties; it does so with causal roles. But we have not quite said what causal role functionalism says that mental properties \textit{are}. Philosophers who wish to adopt causal role functionalism have a choice. If they wish to adopt a \textit{reductionist} view of mental properties, in which mental properties are literally the same thing as -- “nothing over and above” as the expression goes -- neurological properties, then the philosopher may adopt what is called “realizer functionalism.” On this view, mental properties are \textit{identified} with the neurological properties that turn out to play the causal roles provided by the theory $T$. This was Lewis’s own way of running the view. Famously, he used causal role functionalism as an argument for reductionism. The argument went like this:

\begin{quote}
(L1) Mental state $M = \text{the occupant of causal role } R$ (by definition of $M$).

(L2) Neural state $N = \text{the occupant of causal role } R$ (by the physiological theory).

(L3) $\therefore$ Mental state $M = \text{neural state } N$ (by transitivity of $=$).\textsuperscript{15}
\end{quote}

Since the mental property is characterized by a causal role (since, intuitively, the mental property is the thing that \textit{has} the “syndrome of most typical causes and effects”) it must be the occupant of the causal role. But the thing that participates in actual causal transactions is a

\textsuperscript{14} See section IV of Lewis’s article “How to Define Theoretical Terms,” the section titled “The Definitions of T-Terms” (Lewis, 1983b, pp. 87-88).

\textsuperscript{15} Everything but the “(L1)” “(L2)” and “(L3)” is from Lewis (1999a, pp. 248-249).
physical thing, a neurological property, so it must be the occupant of the role, too. But that means that the two properties must be the same property.\textsuperscript{16}

On the other hand, philosophers who wish to adopt a non-reductive view of mental properties may adopt the non-reductive version of causal role functionalism. According to non-reductive causal role functionalism mental properties are not identical to the neurological properties that occupy their definitive causal roles. Rather, mental properties are second-order properties describable as \textit{the property of having a property that plays the relevant causal role.} Such second-order properties are also called “role properties.” Here is a more precise definition of a role property.

\begin{equation}
\text{(GRP) Role property } R \equiv \text{the property of having some property or other } P \text{ play causal role } C.
\end{equation}

Here is how this works. Suppose that \( T \) defines terms for each mental state of which humans are capable, and suppose there are \( n \) many such terms. Suppose “pain” is the \( j \)th term of the \( n \) terms that are defined by the theory. For an entity \( E \) to satisfy \( T \) it must have \( n \) physical properties (in the case of humans these will be neural properties) that interact with each other and with inputs and outputs as \( T \) describes. Then, for each mental term defined by a causal role definition derivable from \( T \), there will be a physical property \( p \) of \( E \) that corresponds to it. In the case of the \( j \)th term, “pain,” the physical property might be c-fiber firing. But “pain” does not refer to c-fiber firing, according to non-reductive functionalists. Rather, “pain” refers to the property of having the \( j \)th role played, the \( j \)th term satisfied by \( E \).

\textsuperscript{16} Lewis was not painstakingly clear over the course of the papers in the late 1960s and early 70s in which he developed his view whether he meant that the occupants of causal roles were properties or property instances -- aka “states.” In his 1966 paper he specifically says that he is interested in universals (Lewis, 1983a, p. 99, f.1), which suggests properties. But universals as such don’t cause things, their instances do and if we take “states” to be co-referential with “instances” then he is referring to instances in the later paper, as quoted above. This sort of issue -- whether the theory is about universals or instances of universals -- will become important later on, when Kim develops his critique of non-reductive causal role functionalism in the 1990s.
Here is a small problem, the solution to which will help us to understand how causal role functionalism works. Recall that each causal role definition is merely a rewrite of the entire theory, with the definiendum moved out front. This evidently means that for entity E to satisfy any term in T it must satisfy every term in T. So how does the non-reductive functionalist propose to distinguish the property of having the jth role played from the property of having the kth role played? Why isn’t every role property identical to the property of having the entire theory T satisfied?

The solution to this worry is to observe that there is a difference between being able to have one’s c-fibers firing and having one’s c-fibers firing. The former is enough to contribute to the satisfaction of T while the latter is what is needed to have a specific role defined by T played. This suggests that the causal role players must be \textit{occurrent properties}, properties had by the subject in on-again, off-again ways. Thus, Sydney Shoemaker:

\begin{quote}
One can think of the core realizer as a property whose instantiation comes and goes as the instantiation of the realized property comes and goes, while the non-core part of the total realizer (what we might call the “surround”) is a relatively permanent part of the subject. (Shoemaker, 2007, p. 21)
\end{quote}

Core realizers are causal role-players. A surround is a property of \textit{being able to have} a specific n-tuple of properties that satisfies T. None of the members of the n-tuple need to be “switched on” for T to be satisfied (= for the surround to be had), but the relevant member does need to be “switched on” -- \textit{occurrent} -- for a given role to be played (= for a mental property to be had). A \textit{total realizer} is a conjunctive property whose conjuncts are a core realizer and a surround -- both must be had in order unconditionally to guarantee that a role property is had. The difference between core and total realizers is what keeps every role property definition from
being satisfied by the same property of an entity that satisfies $T$: that entity’s surround. Just having a surround *would be enough* to satisfy every t-term in $T$ and thus to have every mental property at once if core realizers did not need to be occurrent properties.

But in the last few paragraphs I have introduced a lot of new terms, such as “core” and “total realizer.” Let us step back, slow down, and look at this again.

Over the next few sections (§5-8) I will discuss some key moments in the development of causal role functionalism after its invention by Lewis. In §5 I will discuss the introduction by Ned Block of the idea that causal role functionalism has reductive and non-reductive variants. In §6 I discuss the introduction, by Sydney Shoemaker, of the core/total realizer distinction. These two sections are about developments in the philosophy of mind at the beginning of the 1980s. Then, in the next two sections, §7 and §8, I discuss two arguments that Jaegwon Kim introduced into the philosophy of mind in the 1990s. These arguments set the agenda for work in the metaphysics of mind to the present day. One way to think about Kim’s arguments is to take them as questioning the viability of Block’s work: of the non-reductive variant of causal role functionalism that Block introduced. In contrast, much of my own work in this dissertation can be understood as criticizing Kim for not appreciating Shoemaker’s work: the distinction between core and total realizers. (Though my thesis is actually much more general than this way of putting it might make it seem. In §9 I will make an initial presentation of how my thesis works.)

§5

David Lewis more or less invented causal role functionalism in a series of papers in the late 1960s and early 1970s (Lewis, 1983a, 1983b, 1983c, 1999a). The next significant
developments in the story of causal role functionalism occurred in 1980 and 1981 with the
publication of Ned Block’s “What is Functionalism?” and Sydney Shoemaker’s “Some Varieties
of Functionalism,” respectively (Block, 1980; Shoemaker, 1981). I will discuss Block’s
contribution in this section and Shoemaker’s in the next.

Block showed how Putnam’s machine tables could be translated into Ramsey-Lewis
sentences of the sort in which theories such as $T$ are written. Anything that one wanted to do
with Putnam’s machine table functionalism, then, one could do with Lewis’s causal role
functionalism (Block, 1980, pp. 173-174). But since Putnam’s machine table functionalism was
a non-reductive version of functionalism (machine tables are not identical to their instantiations),
what Block in effect did was show how to give Lewis’s causal role functionalism a non-reductive
variant. Block called the non-reductive variant of causal role functionalism “the functional state
identity theory” since it identifies mental states with what Putnam called the abstract functional
or “machine” states of an entity. (In later years “functional states” would be called “role
properties.”) Block called Lewis’s original, reductive variant of causal role functionalism the
“functional specifier view” since according to this view causal roles specify the neural properties
that are mental properties (Block, 1980, p. 179). In later years the non-reductive version of
causal role functionalism would more commonly be called “role functionalism” and the
reductive version of causal role functionalism would be more commonly called “realizer
functionalism.” Here is how the distinction is described in a recent paper by Don Ross and
David Spurrett:

This distinction (which is stated and clarified in slightly different
terms in [(Block, 1980)]) also allows two importantly different
ways of being a functionalist. The difference turns on whether one
is inclined to identify the functional states with the role they play,
or with what is the realizer of that role in a given case. Put another way, a functionalist might think that “pain” or “money” pick out either the property of having some other (physical) property that realizes pain or money, or that, properly analyzed, they pick out C-fibers firing or dollar bills. Saying that the description of the functional state picks out the role indicates commitment to the view that even though their realizers could be very different, humans and, say, Martians could be in the same mental state when in pain. On the other hand, tying the function to the realizer entails that humans in pain and Martians in pain are in different states, perhaps different types of pain, just because the realizers of the roles in each case are different. (Ross & Spurrett, 2004, p. 605; italics original)

Non-reductive functionalists can equally well call “role properties,” “realized properties.” The fact that these two names for the same kind of unreduced property are about equally common in the literature is an indication of the convoluted history of functionalism. Putnam, as we saw, brought talk of “realization” and “realizers” into the philosophy of mind with his papers in 1960 and 1967 (from, I hypothesize, a previous, more general use of such terms as evidenced in Nagel’s address to the APA.) Lewis at first -- in his 1966 paper -- made no use of the terms “realize” or “realizer”; he just spoke of “causal roles.” But after reviewing Putnam’s 1967 paper (Lewis, 1969) Lewis started using the term “realize” in his own work, though Lewis spoke of the realization of theories, not states or properties (Lewis, 1983b, 1999a). Some commentators have found great significance in the difference between Putnam’s talk of the realization of states and Lewis’s talk of the realization of theories, referring to them as different kinds of realization, material and linguistic (Endicott, forthcoming; Gillett, forthcoming). No doubt there is something to that, but I suspect that as a matter of philosophical history the source of the difference is no deeper than this: Lewis wanted to talk of realization after seeing Putnam do it,
but Lewis had theories and a reductive view rather than machine tables and a non-reductive view, so theories were the only things Lewis could speak of as being “realized.”\footnote{It is sometimes helpful to say of reductive functionalists that they identify realized properties with their realizers. On this way of talking, “realized property” does not mean “role property,” it simply means, “the target property that the functionalist theory is trying to account for.” Thus a reductive functionalist about mental properties may call mental properties “realized” but also say that mental properties are identical to their realizers, and therefore that realized properties are identical to their realizers. I trust that the two ways of taking the term “realized property” are not too confusing, and in any case familiar to most readers.}

Given all of this, it seems correct to say that the idea of a “role property” -- a property that features in causal role functionalism and is not identical to the base property that plays the role, was introduced by Block. This is another way of describing how Block showed that causal role functionalism has the resources for a non-reductive variant. There are two further points to make about role properties, before I move on to Shoemaker’s contribution to the understanding of causal role functionalism.

First, though I claim that Block introduced the idea of an unreduced role property as an element of causal role functionalism -- implying that Lewis did not earlier introduce it -- Lewis later accepted the idea that causal role functionalism implies the existence of causal role properties without particularly allowing that Block had introduced them. Rather, Lewis wrote that the reductivist view -- his own view -- simply involves the denial that role properties are the correct properties with which to identify mental properties. Writing in 1994, he asserted that role properties are disjunctive properties -- by which he presumably meant that role properties are identical to the disjunctions of all of their possible role-playing base properties -- and therefore cannot be causally efficacious. This makes it seem as though reductive and non-reductive functionalism are basically the same theory, with the dispute between them being over which sort of property (of the sorts of property that are provided by causal role functionalism) is the correct sort with which to identify mental properties. The multiplicity of options, I am saying, is Block’s
innovation, though Lewis seems to read it back into his own earlier work. Here is Lewis in 1994, giving a host of reasons why role properties cannot be mental properties:

[The role property] is not the occupant of the M-role. It cannot occupy that or any other causal role because it is excessively disjunctive, and therefore no events are essentially havings of it. To admit it as causally efficacious would be lead to absurd double-counting of causes . . .

Since the highly disjunctive property of being in M does not occupy the M-role, I say it cannot be the referent of M. Many disagree. They would like it if M turned out to be a rigid designator of a property common to all who are in M. So the property I call ‘being in M’ they call simply M; and the property that I call M, the occupant of the M-role, they call ‘the realisation of M’. They have made the wrong choice, since it is absurd to deny that M itself is causally efficacious. (Lewis, 1999b, p. 307)

Lewis here says a number of questionable things. For example, philosophers who adhere to non-reductive causal role functionalism do not deny that neural realizer properties are what occupy causal roles. They don’t deny that neural properties cause the effects of mental properties. What non-reductivists tend to assert is that while it is true that neural properties (properties that count as Ps in GRP) cause the effects of mental properties, role properties (properties that count as Rs in GRP) also cause the effects of mental properties; further, mental properties are role properties. (Thus, pace Lewis, non-reductivists don’t deny that mental properties such as M are causally efficacious.)

Non-reductive functionalists bite the bullet; they try to get away with what Lewis described as “absurd double-counting of causes.”18 This affirmation of double causation is sometimes put as the less-than-happy claim that both the realizer and realized property occupy or

---

18 This generates the problem of causal exclusion, which has been the topic of years of debate in functionalist philosophy of mind. I discuss this problem in chapter three.
play the causal role in question.\textsuperscript{19} Non-reductivists \textit{may or may not} agree with Lewis that role properties are “disjunctive” properties; if they agree, then they may deny \textit{either} that the disjunctions involved are “excessive” \textit{or} that excessive disjunction is a debilitating problem for the causal efficacy of the realized properties.\textsuperscript{20} If they don’t agree that role properties are disjunctive, they are obligated to say that role properties are something other than disjunctions of all of a given causal role’s possible players. There are a lot of ways to read GRP; a point I develop in chapter four.

\section*{§6}

In his 1981 paper “Some Varieties of Functionalism” Sydney Shoemaker introduced one of the most important distinctions in causal role functionalism, the distinction between \textit{core} and \textit{total} realizers. A \textit{core} realizer is a causal role-player. It is a member of an n-tuple of base properties that satisfy a theory that provides causal roles; in the case of the theory $T$ the base the roles are psychological roles and the properties might be neural properties. Core realizers are the base properties that have the syndromes of most typical causes and effects of a mental state; they are properties of the same sort as the neural state $N$ of Lewis’s L2. On Lewis’s version of reductive functionalism, mental properties \textit{just are} core realizers.\textsuperscript{21}

\footnote{Shoemaker uses this as a criticism of standard non-reductive causal role functionalism and a reason to prefer his own “subset” version according which all properties are sets of “causal powers” and realized properties are simply subsets of the causal powers of their realizers. This allegedly makes the duality of role players less problematic (Shoemaker, 2007, p. 5).}

\footnote{The former route is generally taken to be more appealing than the latter. Philosophers who argue that role properties are disjunctive but not excessively disjunctive include Block (1997), Antony (2003) and Clapp (2001). I interpret Gene Witmer as investigating the latter option -- that excessive disjunction is not a problem -- in Witmer (2003). More on these matters in chapter four.}

\footnote{At least, until Lewis published “Mad Pain and Martian Pain,” in which he complicated matters a bit (Lewis, 1983c). Still, at no point did Lewis identify mental properties with \textit{total} realizers, as would Kim in the 1990s.}
It is a familiar and important fact about functionalism in the philosophy of mind that it does not hold that just because a creature has a core realizer that the creature thereby has a mental state. To give the standard example, the core realizer for pain in humans is c-fiber firing; functionalists do not hold that a person has pain just in case that person has c-fiber firing. Nor do functionalists hold that a c-fiber floating in outer space firing all on its own instantiates an instance of pain. Reductive and non-reductive functionalists will agree on these points. Rather, according to both versions of functionalism, c-fiber firing firing must, when it occurs, be playing the causal role of pain in a system, in order for an instance of pain to be had (in which case both the property of c-fiber firing and the property of pain are had by the system, not the c-fiber).

This is easiest to see if we think in terms of property instances, considered as ordered pairs. Here are two ordered pairs, each consisting of an object and a property, that might be meant by term “property instance of c-fiber firing”: (i) <c-fiber, firing>, (ii) <brain, c-fiber firing>. The second one is the one that realizes an instance of pain in Smith, not the first, according to the sorts of realization we are considering. (In chapter five we will look at versions of functionalism that might say it is the first.)

The c-fiber must be, as is said, “hooked up” in the right ways in a system that allows the c-fiber to play the causal role of pain, allows the firing of c-fibers to result in and from the syndrome of most typical causes and effects of pain.

*Total* realizers, on the other hand, are more complicated conjunctive properties. For an entity to have a total realizer of pain is for it to have a core realizer of pain and a “surround” that is appropriate to the core. Let us take as an example Smith, a typical human who has just mildly stubbed her toe. Smith has an n-tuple of neural properties at the ready to fire in causal patterns.
as described by T; thus she has a particular n-tuple of neural properties that satisfy T. The painth member of this n-tuple is c-fiber firing. Smith’s property of having this n-tuple is her surround. C-fiber firing is the core realizer, for Smith, of pain. Once she stubs her toe, her c-fiber fires and she thereby has the property of having her c-fibers firing. At the moment she stubs her toe, then, she has the conjunctive property of having the surround and the painth core. Thus she has at that moment a total realizer for pain.22

Here are some key points about core and total realizers. First, total realizers are unconditionally sufficient for the realized properties that they realize. Core realizers are not. Realized properties thus strongly supervene upon total, not core, realizers. Here is a definition of strong supervenience, taken from Kim:

Supervenience. Mental properties strongly supervene on physical/biological properties. That is, if any system s instantiates a mental property M at t, there necessarily exists a physical property P such that s instantiates P at t, and necessarily anything instantiating P at any time instantiates M at that time. (Kim, 2005, p. 33)

If physicalism and causal role functionalism are both true, then for every entity and every mental property, if the entity has a mental property then the entity has a physical total realizer

---

22 It’s right around here that we can start to see that the relationship between a core realizer and the property that it realizes in canonical cases of causal role realization is not like cases that are sometimes put forward as examples. A typical such example is the property of being a bottle stopper. We may say that an object has the property of being a piece of cork with such and so dimensions, and that these properties allow it to play the causal role of stopping wine from leaving bottles with openings of such an so a shape. I have no problem with saying that in the generic sense of “make real” the property P of being a piece of cork (had by object C) with such and so dimensions realizes the property B of C of being a bottle stopper, but I deny that this sense of realize is the sense that is explicated via the canonical sense of causal role playing that I am describing. This is so for several reasons.

(i) P does not realize B because P or C are playing a role in any specific system. C is just a bottle stopper, for any appropriate bottle/system. There is no sense in which a c-fiber just is a pain, for any appropriate system.

(ii) P is the core realizer for B; P does not require that there be an “active surround” in order to realize B. This is primarily a restatement of (i).

(iii) Even when C is actively playing a part in a system, i.e. stopping a bottle, it’s “surround” is very different than the type of surround discussed by Shoemaker in connection with canonical causal role functionalism. A surround is a property of being able to have but not necessarily having a specific n-tuple of properties. On the other hand, in order for C to play an active role in a system all the objects (the bottle, the wine) must in fact have their respective realizer properties.

(iv) P is a permanent property of C; therefore so is B. In canonical cases of causal role playing, core realizers are occurrent, occasional properties, as are realized properties.
such that any entity that has that total realizer will have that mental property. No parallel claim is intended or defensible concerning core realizers.

Second, the distinction between core and total realizers must be respected by both reductive and non-reductive functionalists. Since no functionalist thinks that a core realizer all by itself can be sufficient for a realized property, no reductive functionalist ought to think that realized properties are necessarily identical to core realizers -- and as far as I know, none do.\(^\text{23}\) Reductive functionalists are free to think that mental properties (here I speak of properties, not property instances) are identical to total realizers, since there is no problem with a property being identical to a property that is unconditionally sufficient for it. Of course, such a view requires the denial of multiple realizability -- a denial that a mental property could have multiple non-identical total realizers.

If one accepts that a given realized property has multiple possible realizers then one may adopt the view that the realized property is identical to the disjunction of all of the possible total realizers. Though Jaegwon Kim would present this idea in a 1992 paper (though he left implicit at best the required restriction to total realizers) it was in fact suggested by Shoemaker in the 1981 article we have been discussing. Indeed, though Kim will later say that it is obviously a reductive move to accept that role properties are identical to disjunctions of their realizers, Shoemaker will advance a more nuanced idea; an idea about the reductive status of disjunctive role properties that has popped up from time to time in the literature but has not gained wide

\(^{23}\) Lewis held that mental terms referred to core realizers, but he did not think that the mental terms were rigid designators. Thus it would be incorrect to say that he held that mental properties were necessarily identical to core realizers. To hold Lewis’s view one must hold a description theory of names according to which names like “pain” are hidden definite descriptions that may be satisfied by now this, now that core realizer.
A natural suggestion is that a functional property can be identified with the property which is the disjunction (perhaps infinite) of all of its total realizations. Now it may be that all possible total realizations of mental properties (or of functional properties generally) are physical properties, and that case each functional property will be identical to a physical property -- assuming that it is permissible to speak of infinite disjunctions. But it cannot be excluded a priori that it is possible that functional states should have realizations that are wholly or partially nonphysical (perhaps involving, as Putnam once suggested, “bundles of ectoplasm”). And if that is a possibility, then the disjunctive property with which such a functional state is identifiable will have some nonphysical disjuncts, and so cannot be said to be a physical property. But even if this is so, if it is nevertheless true that only the physical disjuncts of such properties are ever instantiated in the actual world -- or, in other words, if all actual realizations of functional properties are physical -- then materialism will be true. If one interprets materialism as requiring that all properties be identical to physical properties, and if one also allows for uninstantiated properties and holds that the disjunction of any two properties is a property, one is in effect taking materialism to hold, not only that there are no nonphysical entities, but that it is logically impossible that there should be any. And that, surely, is a much stronger claim than the materialist is committed to simply in virtue of being a materialist. (Shoemaker, 1981, p. 98)

Though Shoemaker’s distinction between core and total realizers has for the most part been remembered and put to use, the ideas in this passage made no impact. They had to be rediscovered piecemeal later on. To this day the group of philosophers that I call the “non-reductive disjunctivists” -- we’ll get to them soon -- typically assume that mental properties are identical to the disjunction of only their physically possible realizers, without apparently being aware of the options Shoemaker provided in 1981. They therefore cut themselves off from an

---

obvious response to the first of two arguments against non-reductive causal role functionalism that Kim put forward in the 1990s. It is to Kim’s two arguments, and the way that they shaped functionalist debate in the 1990s, that we now turn.

§7

In this section and the next I present two arguments against non-reductive functionalism presented by Jaegwon Kim in the 1990s: the Disjunction Argument and the Supervenience Argument, along with some concepts that Kim introduced to make these arguments work. I will discuss the Disjunction Argument in this section and the Suprvension Argument in the next. These discussions will be fairly cursory: a much more extensive discussion of the Disjunction Argument may be found in chapter four; of the Supervenience Argument, in chapter three. For now I am interested in recounting some history: the basics of the arguments, what opponents said in response, and how these moves and counter-moves shaped and limited the debate over the status of role-properties in the 1990s.

Here is how Kim’s Disjunction Argument works. Take as an example the mental property of being in pain. Assume, with the non-reductive causal role functionalists, that it is identical to a role property of the sort defined by GRP. Assume that the role property pain has three non-identical possible total realizers: N1, N2, N3. Stipulation: the property of having a one of a list of properties is identical to the disjunctive property whose disjuncts are the members of the list. It follows that pain is identical to the disjunctive property named by the predicate “N1 v N2 v N3.” But N1, N2 and N3 are non-identical properties by hypothesis, so by the Principle of
Causal Individuation of Kinds they must have different causal powers.\textsuperscript{25} But by the Causal Inheritance Principle that means that the role property must have different causal powers on different occasions.\textsuperscript{26} But that means it is unprojectible, and therefore not scientifically legitimate. On the other hand, if N1, N2 and N3 aren’t so different after all, if their disjunction is causally tidy enough to count as projectible, then the disjunction is legitimate but nothing over and above what N1, N2 and N3 have in common. So the role property is either illegitimate or trivially reducible to its base. In either case a reductive conclusion has followed from allegedly non-reductive premises.

Kim’s Disjunction Argument persuaded most non-reductive functionalists that they had to accept that role properties probably were identical to disjunctions of their realizers. This was in part because most non-reductive functionalists accepted the \textit{Causal Inheritance Principle} upon which the Disjunction Argument rests (see footnote 25). In response to Kim’s argument most non-reductive functionalists tried to find ways to argue that identifying mental properties with disjunctions of realizers did not or need not count as a reductive move. They did this for the most part without remembering or anyway without mentioning the distinction, suggested by Shoemaker, between disjunctions of metaphysically possible realizers and disjunctions of physically possible realizers. Like Kim, the non-reductivists for the most part took it that the

\textsuperscript{25} Kim:

\begin{quote}
[Principle of Causal Individuation of Kinds] Kinds in science are individuated on the basis of causal powers; that is, objects and events fall under a kind, or share in a property, insofar as they have similar causal powers. (Kim, 1993, p. 326)
\end{quote}

\textsuperscript{26} Kim:

\begin{quote}
[The Causal Inheritance Principle] If mental property $M$ is realized in a system at $t$ in virtue of physical realization base $P$, the causal powers of this instance of $M$ are identical with the causal powers of $P$. (Kim, 1993, p. 326; italics original)
\end{quote}
battle had to be waged over the status of disjunctions of realizers with only physical disjuncts. So they had to find other ways of denying the reductive nature of the disjunctive move.

What none of the non-reductivists did was question the Causal Inheritance Principle itself. I believe that this is because it was widely assumed that the Causal Inheritance Principle was, for all intents and purposes, merely another way of stating how causal role realization works. If N1 is making real an instance of M by playing the causal role of M then the instance of M can do no causal work except inasmuch as N1 is doing it too, the thought would go. So if on different occasions N1, N2, and N3 play the role of M and N1, N2, and N3 do different kinds of causal work then on those various occasions the various instances of M will do different kinds of work. On the other hand if N1, N2, and N3 do very similar kinds of work then one might as well say that M is just what N1, N2, and N3 have in common and so M is after all nothing irreducibly “other” than they.

As it happens, the non-reductivists were more concerned about the first horn of Kim’s dilemma than the second. It concerned them more that M might turn out not to be a “natural kind” than that it might turn out to be a natural kind with a lot in common with its realizers. So the non-reductivists tried to show how the physical realizers of a given mental property probably really do have causal powers in common and therefore M is not illegitimate. From there the non-reductivists tried to learn to live with the less ambitious versions of non-reductivism that were left to be defended. At least M isn’t identical to any one of the disjuncts; that’s a sort of non-reductive result.

Again, I want to stress the peculiarity of not noticing the option Shoemaker had given the non-reductivists back in 1981. By the lights of the later debate over the Disjunction Argument,
Shoemaker had crafted an option that we might call “semi-reductive semi-materialism” according to which a mental property is identical to a causally homogenous disjunction of properties, not all of which are physically possible. I also want to stress the peculiarity of not even attempting to attack the assumptions and stipulations that I mentioned in my brief rehearsal of the argument: that the property of having a disjunction is identical to the disjunction; that the property of having one of a list of properties is identical to the property of having the disjunction of the members of the list; the two Principles. These were all accepted by most of the non-reductivists in the debate (the lone rouge being, unsurprisingly, Jerry Fodor, in his (1997)).

§8

Kim’s Supervenience Argument has had, if anything, even more influence on the debate over the reductive or non-reductive status of causal role functionalism than the Disjunction Argument. I will describe the workings of the Supervenience Argument in detail in chapter three. For now I will just give a brief description of it.

The Supervenience argument elaborates on the “absurd double-counting of causes” mentioned by Lewis; indeed the problem exploited by Kim for this argument was alluded to as far back as Nagel’s address to the APA, when Nagel mentioned that realized qualities don’t participate in the realization of anything else.

Take a typical example of mental-mental causation. Smith’s thought about cake (M1) causes Smith to have a desire for cake (M2). M1 and M2 must both have neural properties as strong supervenience bases; call these neural properties N1 and N2 respectively. We want to be able to say that Smith has M2 because Smith had M1. But given the M2 strongly supervenes on
N2, N2 is a completely sufficient explanation for the existence of M2. We don’t need to bring M1 into the explanatory picture at all. Suppose, to save the relevance of M1, we try saying that it causes M2. But M1 has a supervenience base in N1, and presumably N1 provides a sufficient causal explanation for the existence of N2. So there is nothing left for M1 to do. It has been excluded from the causal explanation. The only way to save the relevance of M1 is to identify it with N1. But that is the reductive conclusion that Kim wants and his opponents to not.

In response to this Argument the non-reductivists tried various responses. I describe them in chapter 3. For now what I want to stress is that all of the responses that the non-reductive causal role functionalists have given to the Supervenience Argument have accepted an unstated principle upon which the Supervenience relies. Unlike the Principle of Causal Individuation of Kinds and the Causal Inheritance Principle, Kim never set this one out and gave it a name; very likely he believed it was too obvious to bear pointing out. Nevertheless I will do so:

_The Principle of Causal Subveners_ For any instance of a realized property R, the realizer property P upon which the instance of R strongly supervenes figures in the causal explanation of the effects of that instance of R.

I deny the Principle of Causal Subveners. This denial and my defense of it is one of the primary contributions of this dissertation. I argue in chapter two that the properties upon which role properties strongly supervene -- that is to say, total realizers -- don’t figure in causal explanations at all.
Over the next two sections I will lay out the contents of parts II and III of this dissertation. In this section I will explain how Part II of the dissertation works. I explain how each chapter of Part II works individually as well as how they work together to defend the premises of the master argument of this dissertation, which I also present in this section. In §10 I will say something about Part III.

Part II has three chapters. Together they comprise a single argument against the viability of any form of causal role functionalism, reductive or non-reductive, in the philosophy of mind or indeed for any special science.

Part II begins with Chapter Two; there, I present an argument for the denial of the Principle of Causal Subveners. This denial has, if correct, quite broad and fairly drastic consequences for current views and concerns in the philosophy of mind. In Chapter Three presents an example of this, as I show how the falsity of the Principle of Causal Subveners renders Kim’s famous Supervenience Argument straightforwardly unsound. This may seem like good news for the non-reductive versions of causal role functionalism that the Supervenience Argument is meant to attack; but it is not, since the falsity of the Principle of Causal Subveners renders such non-reductive views just as problematic as the reductive views that the Supervenience Argument is supposed to motivate.

In Chapter Four I present another problem for non-reductive versions of causal role functionalism. I argue that the notion of a role property is deeply unclear, because GRP defines at least two kinds of property. This in itself is an innovation; not nearly enough attention has been given to the multiplicity of notions of role property that do work in discussions of non-
reductive physicalism; non-reductive causal role functionalists have been woefully slow to articulate just what kind of property they think GRP defines when they are articulating and defending their views. This kind of unclarity has made it very difficult to see just what is going on when arguments over the status of causal role functionalism that rely on the exact nature of role properties are put forward. Another of the points of Chapter Four is to critique Jaegwon Kim’s other important argument from the 1990s, the Disjunction Argument, as well as responses to it. But the most important goal of Chapter Four is to make use of the critique of the Disjunction Argument as an occasion to demonstrate that GRP is ambiguous, and that none of the disambiguations of GRP are particularly attractive as candidates for identification with realized properties, for example mental properties.

The three chapters of Part II are to be taken as a single extended argument. Or, better, they should be taken as defenses of the premises of a single argument, which I present now. Here is the argument:

(1) In order for causal role realization as employed in causal role functionalism to be a viable explication of the notion of the realization relation, it must be the case that realized properties are identical to one of the kinds of property that figure in causal role realization. Those kinds of property are: (i) core realizers, (ii) total realizers, (iii) role properties considered as disjunctions of total realizers, (iv) role properties considered as properties of having a role played.

(2) Realized properties such as mental properties are causally efficacious. (Denial of epiphenomenalism.)

(3) For any two properties x and y, x is identical to y only if x strongly supervenes on y. (Trivial.)

(4) Total realizers are not causally efficacious. (Chapter 2.)

(5) Disjunctions of causally ineficacious properties are not causally efficacious. (Obvious.)
(6) Disjunctions of total realizers are not causally efficacious. ((4,5); Chapter 4.)

(7) Realized properties do not strongly supervene on core realizers. (Definition of “core realizer.”)

(8) Realized properties are not identical to (i) core realizers. (3, 7)

(9) Realized properties are not identical to (ii) total realizers. (2, 4)

(10) Realized properties are not identical to (iii) role properties considered as disjunctions of total realizers. (2, 6)

(11) Role properties of the sort mentioned in (1), (iv) are not causally efficacious. (Chapter 4.)

(12) Realized properties are not identical to role properties of the sort mentioned in (1), (iv). (2, 11)

(13) Realized properties are not identical to any of (i-iv): core realizers, total realizers, role properties considered as disjunctions of total realizers, or role properties of the sort mentioned in (1), (iv). (8, 9, 10, 12)

(14) Causal role functionalism is not a viable explication of the notion of realization. (1, 13)

Despite its length, the argument is straightforward, and straightforwardly valid. In order for any version of causal role functionalism, reductive or non-reductive, to be correct in the philosophy of mind, it must be the case that mental properties are identical to one of the kinds of properties that figure in accounts of causal role functionalism. Those kinds of properties are: core realizers, total realizers, role properties considered as disjunctions of total realizers, role properties considered as something else. The thesis of Part Two of my dissertation is that mental properties cannot be identical to any of those four.
Part III has only one chapter, though it is a long one: chapter five. Most of the work of introducing chapter five I do at the beginning of chapter five. But I will make some preparatory remarks here.

In §1-9 of this introduction I have discussed a particular explication of the notion of realization, causal role realization. Causal role realization is by far the most common way of explicating the notion of realization, at least when it comes to one broad category of alleged cases of realization: cases in which the realized property and its realizer property are both had by the same object. In such cases it is most common to say that the realized property is a role property: the property of having some other property that plays a specified causal role. In order for an object to have a role property it must, by definition, have the property that plays the role.

But some philosophers believe that there are cases of realization in which the realizer property or properties are had by entities other than the object that has the realized property. In such cases the causal role-playing relation is not even a candidate for being the correct relation with which to explicate the notion of realization. So philosophers who apply the notion of realization to such cases are obligated to come up with some other way of explicating the notion of realization.

It will be useful to have terms for these broad categories of realization. I give the name “same level realization” to forms of realization, be they reductive or non-reductive, in which one and the same object is supposed to have the realizer and realized properties. I give the name “cross-level realization” to forms of realization in which the objects that have the realizer properties are not the same as the object that has the realized property. I call the latter of these
two forms of realization “cross-level” because philosophers who employ the notion generally have in mind cases in which the objects that have the realizer properties are *proper parts* of the object that has the realized property. Further, it is usually taken to be the case that the proper parts of the object exist at a lower level of nature than the object that has the realized property. (For current purposes, to say that one object exists at a lower level of nature than another is to say either that the former is studied by a more fundamental special science than is the latter, or that the former is studied by fundamental physics and the latter is studied by a special science.27)

It is possible to be a reductivist or a non-reductivist about either same- or cross-level realization. (Recall that a reductivist about realization is a philosopher who holds that the realized and realizer properties are, in the end identical; a non-reductivist about realization denies this.) With this in mind we may divide accounts of realization into four broad categories along two axes: same-level non-reductive; same-level reductive; cross-level non-reductive; cross-level reductive. We may make a chart of these, as follows -- filling the chart in with prominent proponents of each broad category.

<table>
<thead>
<tr>
<th></th>
<th>Non-Reductive Realization</th>
<th>Reductive Realization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Same-Level Realization</strong></td>
<td>Putnam, Antony, Levine</td>
<td>Lewis, Kim</td>
</tr>
<tr>
<td><strong>Cross-Level Realization</strong></td>
<td>Gillett, Lycan</td>
<td>Heil</td>
</tr>
</tbody>
</table>

Table 1: Broad Categories of Realization

So here is another way to think about the structure of this dissertation. Part II of this dissertation is about the upper row; Part III is about the lower row.

---

27 It would be more precise to hold that it is properties, not objects, that are the targets of study of the sciences. Then a property P falls under the purview of a science Sc just in case the laws in which the predicate that refers to P feature are laws studied by Sc. This makes the mereological conception of levels problematic, however, and to no good current purpose. So I shall stick with the mereological conception of levels discussed in the text.
The exception to this schema is that I have very little to say about the box in the lower right of Table 1, cross-level reductive realization. Cross-level reductive realization, while taking up a legitimate spot in the logical space of available options, requires a very unusual way of unpacking the notion of realization -- as far as I know, John Heil is the only prominent philosopher who attempts it (Heil, 2003). To see why it requires some odd maneuvering, note that it requires identifying the realized property instances of an object with property instances had by that object’s proper parts. The only way to make that work is to be eliminative, in at least some sense of “eliminative,” of both the realized properties and the object that supposedly has them. Basically, one must construe “x realizes y” as meaning “the predicate ‘y’ indirectly refers to and is made true by instances of x, even though instances of x occur at lower levels of nature than seem to be referred to by ‘y’.” Thus cross-level reductive realization requires a nominalist understanding of the alleged realized properties. (And for that matter a nominalist understanding of the alleged objects that have the alleged realized properties.) Therefore on this way of running things there are no realized properties.

Section three, then, shall be devoted to cross-level non-reductive realization. Rather than focus on one attempt to explicate the notion of cross-level realization, as I do for same-level realization in Part II, in Part III I look at three different attempted explications. I do this because there is no standard or most popular way to make sense of cross-level realization. As I will show, there is not even agreement among advocates of cross-level realization as what the explanatory point of positing cases of cross-level realization is supposed to be.

My primary thesis about cross-level realization can be stated as a dilemma: to the extent that advocates of cross-level realization try to make it explanatory in the same way that the
causal role-playing version of same-level realization is supposed to be explanatory, advocates of cross-level realization end up turning it into same-level realization. To the extent that advocates of cross-level realization manage to keep the cross-level spirit of cross-level realization in play, they render the explanatory point of positing it much more spare than is the explanatory point of positing cases of causal role same-level realization.

This is of special interest because, generally speaking, it is metaphysicians of science of who make use of the notion of cross-level realization, and philosophers of mind who make use of the notion of same-level realization.\(^{28}\) When philosophers of mind posit the realization relation, they are usually positing a relation that fully explains the presence of the realized property. The realization relation, as spoken of in the philosophy of mind, is after all supposed to be the mind/body relation -- the relation that explains how it could be that merely material beings could have mental states. On the other hand, when metaphysicians of science posit a cross-level realization relation, they are obviously not trying to fully explain the presence of the realized property. That task is the task of the sciences themselves. All that the metaphysician of science is trying to do is provide a broad metaphysical framework to capture the various more specific and more explanatory relations and properties that are posited or discovered by the sciences.

To that extent, it makes perfect sense that the explanatory point of positing a same-level realization relation would be different than the explanatory point of positing a cross-level realization relation. But I don’t think that these differences, and the interplay of these different kinds of differences, is well understood. In Part III I try to make some ordering remarks.

\(^{28}\) Thus we get the “two masters” of realization spoken of by Craver and Wilson (Craver & Wilson, 2006).
Finally, in my conclusion, I return to Nagel’s 1954 address to the Eastern Division of the American Philosophical Association. I assess the causal role functionalist project as an attempt to unpack the naturalist worldview Nagel sketched in that talk. It seems to me that causal role functionalism accurately reproduces, or magnifies, or projects, a tension at the heart of the project that Nagel proposed. Nagel’s version of naturalism is able to save everything in our pre-philosophical picture of the world except for the idea that “dreams and joys, aspirations and desires” “figure in their own realization or in the realization of anything else.” That is, Nagel would deny premise (2) in my master argument -- but that premise is exactly the one that that advocates of causal role realization thought that the idea of realization was uniquely suited to save. I think that there probably are ways around this impasse, but to see them we have to set certain core ideas in the philosophy of mind on a new footing. Causal role realization was a product of ideas coming from mid twentieth century philosophy of science and metaphysics. To really get started on the new century, we have to leave it behind.
Part II

Chapter 2: Four Theses About Total Realizers

§1

Causal role functionalism in the philosophy of mind is the view that mental properties are realized properties, and that realized properties are causal role properties. The goal of this chapter is to get clear about the nature of the base properties upon which causal role properties strongly supervene: total realizers. Confusions about total realizers go to the heart of causal role functionalism; getting clear about total realizers is essential to understanding whether and in what ways the causal role functionalist project is viable. This much should be familiar: a total realizer is a conjunctive property whose conjuncts are a core realizer -- also called a causal role-player -- and a much more complex, facilitating property that allows the core realizer to play its causal role. The distinction between core and total realizers is due to a paper from the early 80s by Sydney Shoemaker (Shoemaker, 1981); recently, Shoemaker has returned to the topic of core and total realizers and has given the helpful name “the surround” to the complex facilitating property (Shoemaker, 2007: 21). Thus, a total realizer is a conjunctive property whose conjuncts are a core realizer and a surround.

In the current chapter I wish to establish four theses about total realizers. The first two are these:

(T1) A core’s surround is not the same thing as the core’s causal background conditions.

(T2) Total realizers do not figure in the causal explanations of a core realizer’s effects (or, for that matter, any effects).
(T1) drops out of a correct description of total realizers. I will argue that (T2) follows from (T1). In this paper I will also demonstrate the truth of two further, ancillary theses about total realizers:

(T3) It is possible for two different organisms to have the same core realizer for a given realized property while having different surrounds and therefore different total realizers of that realized property.

(T4) All of a given organism’s psychological total realizers have the same surround.

I take it that T2 is the most obviously significant thesis. Its truth, should it be possible to establish, will have significant consequences for ongoing debates between reductive and non-reductive functionalists in the philosophy of mind, particularly over exclusion and disjunction arguments. But my primary purpose in this chapter is not to draw out the consequences of a correct view of total realizers, but rather, simply to work toward attaining such a correct view.

Establishing all of T1-4 is necessary for that.

I will proceed as follows. In §2 I will give a standard, and standardly truncated, introduction to the notions of core and total realizers. Then in §3 I will provide a more complete and correct account of total realizers, via a brief exegesis of Shoemaker’s originating 1981 article. I proceed in this way -- first offering a bad, and then a better, description of total realizers -- in order to show what I think is misleading about the standard way of describing total realizers, and how to correct it. In §4 I will establish T1. Then in §5 I will establish T2. In §6 I will establish T3. In §7 I will establish T4. Finally in §8 I will discuss the consequences of accepting T1-4. The most serious consequence is that we must acknowledge that total realizers

---

1 I draw out a very big consequence of T2 in the next chapter: that Jaegwon Kim’s Supervenience Argument is unsound.
do not figure in the causal explanations of the effects of realized properties. While total realizers explain causal role-playings, they do not explain causings. A better grasp of the distinction between causal role-playings and causings is one very good reason to get clear about T1-4.²

§2

In this section I will relate a version of what I take to be the common, flawed understanding of the distinction between core and total realizers. Let us begin by reminding ourselves of what total realizers are. Suppose that Smith, a typical human, instantiates some realized property; say the mental property pain. According to the non-reductive version of causal-role functionalism, realized properties such as pain are role-properties -- they are in each case the property of having some property or other that plays the relevant causal role. In the case of pain, pain is the property of having some property or other that plays the causal pain role. (The “pain role” is, to borrow David Lewis’s original phrasing, the syndrome of pain’s most typical causes and effects, e.g. skin damage and wincing.)

Given this, a core realizer of pain is the property had by an entity that plays the pain role for that entity. Smith’s core pain-realizer is the physical or neurological property that she is typically caused to have by the typical causes of pain, and that in turn causes her to have the typical effects of pain. Thus, to over-simplify, since pain is typically caused by skin damage and typically causes wincing, if in Smith the neurological property c-fiber firing is typically caused

² In this chapter and in following chapters I tend to use “causing” and “cause” and “causal explanation” in fairly loose ways. I believe that this is okay -- that I do not need to specify a theory of causation or causal explanation -- because (i) the views that I am discussing are often not clear about causation and also because (ii) my criticisms will apply not matter which specific theories of causation and causal explanation a defender of realization might wish to adopt. As we will see, nothing I say hinges on the specifics of causation or explanation.
by skin damage and typically causes wincing, then \textit{c-fiber firing} plays the causal role of pain in
Smith and (to say the same thing twice) is the core realizer of pain in Smith.

Smith’s \textit{total} realizer of pain, on the other hand, is the core realizer together with
whatever other of her properties make it the case that the core realizer does in fact play the causal
role of pain. Returning to Sydney Shoemaker’s original article on this topic, the additional
properties, considered as one big additional property, are described as “a determinate form of the
property: \textit{being physically constituted in such a way that [c-fiber firing] plays the causal role
definitive of pain}” (Shoemaker, 1981: 75; italics original). Thus, while it is possible for an
organism to have a core realizer but not the corresponding realized property (because the
organism need not be physically constituted in such a way that the core plays the relevant causal
role) it is \textit{not} possible for an organism to have a total realizer but not the corresponding realized
property. This is because, as mentioned, realized properties are causal role properties, and a
causal role property \textit{just is} the property of having a core realizer play the relevant causal role.
Also as mentioned, Shoemaker, in a later work, provides us with a useful name for the additional
big property, the one that facilitates the core’s playing of the role: he calls it “the
surround” (Shoemaker, 2007: 21). Thus, each total realizer is a conjunctive property whose
conjuncts are a core realizer and a surround.

So far so good. What I have done up to now is provide a fairly standard brief
introduction to the notions of core and total realizers. In the next section I will make some
alterations and additions to it. The standard account leaves out some critical points that make it

\begin{footnotesize}
\begin{enumerate}
\item To say that one property \( x \) is a more determinate form of another property \( y \) is just to say that \( x \) is a
determinate of \( y \), where \( y \) is a determinable. Thus \textit{red} is a more determinate form of the property \textit{colored}.
For a more formal account of determinateness see Funkhouser (Funkhouser, 2006). Funkhouser
describes one property \( x \) as a more determinate form of another property \( y \) if \( x \) and \( y \) have the same
determination dimensions and the property space of \( x \) is a proper subset of the property space of \( y \) (554).
\end{enumerate}
\end{footnotesize}
evident, once they are understood, that total realizers do not figure in the causal explanations of their supposed effects.

§3

In making the distinction between core and total realizers Shoemaker appealed to David Lewis’s work on defining theoretical terms (Shoemaker, 1981, Lewis, 1983). As Shoemaker described them, a psychological core realizer is a property that occupies one of the roles provided in the Ramsey-Lewis sentence of the correct psychological theory. On Lewis’s reductivist understanding the theoretical terms defined by the Ramsified psychological theory designate the occupants of the roles. On a non-reductivist understanding, however, the terms designate second-order “role-properties”: properties of having occupants. On either reading, the core realizers are the role occupants. A total realizer is a conjunctive property whose conjuncts are a core realizer and a more complex property: the property of having a specific n-tuple of instantiable (though not necessarily instantiated) core realizers that jointly satisfy the Ramsey-Lewis sentence of the psychological theory. Shoemaker described this complex property (that he would later dub “the surround”) as “a determinate form of the property: being physically constituted in such a way that [the core realizer] plays the causal role definitive of [the realized property]” (Shoemaker, 1981: 97; italics original). But let us back up and go through that again more slowly.

Shoemaker used as an example of a functional property the property pain. “Pain” is a theoretical term, defined by the psychological theory \( T \). On Lewis’s reductive understanding it designates the occupant of the causal role that \( T \) uses to define “pain.” On a non-reductivist understanding, “pain” designates the role property of having some occupant or other that plays
the $T$-defined pain role. $T$ itself is a holistic theory -- it defines all of its theoretical terms at once, using $O$ (pre-theoretical “old”) terms as well as all the other $T$-terms (theoretical terms) to say what any given theoretical term means. This is certainly old hat: what is distinctive about the Ramsey-Lewis method of defining terms is that it defines all of a theory’s terms at once by using a set of mutually constitutive roles.

This means that all of the roles provided by a given theory have to be occupied in order for any of them to be occupied. An organism cannot have a role occupant for pain unless it has a role occupant for every psychological state defined by $T$. This is not, of course, to say that the organism has to have them all instantiated at the same time in order to have any one psychological state instantiated. Rather, the organism simply must be able to have them all instantiated. The roles must all have potential occupants. Were this organism to experience the causes of hunger, then this organism would be caused to have this core realizer of hunger. Were this organism to experience the causes of jealousy, then this organism would be caused to have this core realizer of jealousy. And so on, for each psychological state defined by $T$. To have a surround is to have a specific n-tuple of such role-occupants “at the ready.” To press this point, let me quote at some length from Shoemaker’s original article. Here, Shoemaker abbreviates $c$-fiber firing, the core realizer of pain in an organism, $A$, as “$P_j$.” The core realizers had by $A$ that jointly satisfy $T$ are $P_1 \ldots P_n$. (Of course, $P_j$ is one of $P_1 \ldots P_n$)

“$T(\ldots P_j \ldots )$” will be a general proposition about how states $P_1 \ldots P_n$ relate causally to one another and to inputs and outputs, in whatever creatures these states occur. It will have the same form as the psychological theory from which our Ramsey-sentence was derived; to expose a little more of its form, let us rewrite it as “$\forall x[T(\ldots P_j x \ldots )]$”, where the variable ‘$x$’ takes persons as values. But must this general proposition be true, i.e. must these states behave as it says whenever and wherever they occur, in
order for it to be true that in the case of person A the having of $P_j$ realizes the property of being in pain? . . . On the assumption that being in pain is a functional property derived from theory $T$, the truth of the general proposition $\forall x[T(\ldots P_j x \ldots)]$ should have no bearing on whether A has pain in virtue of having C-fibers firing (i.e., having $P_j$); all that should be required for the latter is that A’s brain be wired so that in him C-fiber firings have the required sorts of causes and effects. What we want as a first conjunct of our physical predicate is not the proposition “$\forall x[T(\ldots P_j x \ldots)]$”, but rather a predicate -- ‘$T(\ldots P_j x \ldots)$’ or ‘Tx’ -- which results from the removal from that proposition of the universal quantifier binding the variable ‘x’; this predicate will be true of a person just in case the person is as all subjects of $P_j$ would have to be in order for “$\forall x[T(\ldots P_j x \ldots)]$” to be true, but it can be true of a person without that general proposition being true. For the same reason, our functional predicate should be rewritten as ‘$\exists F_1 \ldots \exists F_n[T(\ldots F_j x \ldots) & P_j x]$’ or (for short) ‘$\exists F_1 \ldots F_n (Tx & F_j x)$’ and the physical predicate which results from replacing ‘$F_1$’ . . . ‘$F_n$’ with ‘$P_1 \ldots P_n$’ will read ‘$T(\ldots P_j x \ldots) & P_j x$’ or (for short) ‘$Tx & P_j x$’. (Shoemaker, 1981: 96-97)

The total realizer for pain is expressed by the predicate ‘$Tx & P_j x$’ where $P_j x$ is the core realizer for pain and “Tx” is the name Shoemaker gives to the property of having $P_1 \ldots P_n$ relate causally to one another and to inputs and outputs in $T$-appropriate ways. “Tx” refers to a complete and specific $T$-satisfying n-tuple of properties; in other words, a specific surround.

There are two aspects of the above passage that I wish to stress. The first involves the grain of the surround. Shoemaker is at some pains to show that the surround, Tx, is a more determinate property than the property of having any n-tuple of properties that allows $P_j$ to play the pain role. The latter would be, merely, the property of being physically constituted in such a way that $P_j x$ plays the causal role definitive of pain; where what shoemaker wants to define is a more determinate form of that property. Thus, Tx is the property of (potentially) having a specific n-tuple of properties, among them $P_j$, such that when $P_j$ is had it plays the pain role.
Note though that Tx could be more determinate than Shoemaker makes it. Seeing this is crucial to seeing the kind of property that Shoemaker took a surround to be, and why. One could define a property of having a specific n-tuple of physical properties that interact in specific T-appropriate ways. Such a property would specify not just all of the core realizers and the roles they occupy, but also the specific “wiring diagram” of an organism; one that causally “wires” each core realizer to the next. Let us say that in order for a specific ordered n-tuple of properties S that includes P_j and P_m to satisfy T, it is necessary for P_j to cause instantiations of P_m. There are, no doubt, countless biological ways that P_j could be linked to P_m such that instantiations of the former caused instantiations of the latter. The theory T does not discriminate between them -- but it does require that in order for S to satisfy T P_j must play the specific role that it plays and that P_m play the specific role that it plays. Switch them, and a different surround, a surround other than S, results. The reason that Shoemaker did not specify a causal wiring diagram, I believe, is that Ramsey-Lewis sentences don’t provide a way of discriminating between causal wiring diagrams. A surround is the property of having T satisfied by an ordered N-tuple -- two creatures can satisfy T with the same N-tuple while not being otherwise similar, while not having the same wiring diagram. In such a case the creatures have all the same core realizers and in the same roles, and that is as far as the Ramsey-Lewis method probes into their natures. Surrounds are as fine-grained as T is, no more and no less.

Second, consider the holistic way in which T defines psychological terms. A creature must have an n-tuple of core realizers at the ready to satisfy the entire psychological theory, in order for that creature to be correctly describable as having any one of the psychological states mentioned in the theory. Thus, to be capable of pain, a creature must also be capable of anger,
thirst, itch, relief, anxiety, hunger, and so on (and perhaps not just animalistic emotions, if we take the functionalist thesis seriously then it might be that the creature must be capable of the whole human slew of possible psychological states, including jealousy and the belief that \( \{ (a \rightarrow b) \cdot a \rightarrow b \} \).

Accordingly, in order for \textit{c-fiber firing} to play the causal role of pain in Smith it is not enough for \textit{c-fiber firing} to be typically caused in Smith by inter alia skin damage and to cause in Smith inter alia wincing, it is \textit{also} necessary that Smith have some determinate core realizer for \textit{hunger} that is typically caused by inter alia an empty stomach and that typically causes inter alia Smith’s mouth to water in the presence of perceived food. Conversely, in order for some core realizer to play the causal role of hunger in Smith it is not enough that the core realizer be caused in Smith by inter alia an empty stomach and to cause inter alia Smith’s mouth to water, it is \textit{also} necessary for Smith to have some core realizer for pain that is caused inter alia by skin damage and that causes inter alia wincing. And so on for every single one of Smith’s possible psychological states: Smith must have potential core realizers for all of them in order to realize any of them; Smith must satisfy the whole causal/psychological theory in order to realize (in Lewis’s sense of “realize”) any of it.

\textit{But that is not to say} that Smith’s \textit{c-fiber firing} requires \textit{in any sense} some core realizer for \textit{hunger}, in order to cause Smith’s orbicularis occuli muscles to flex -- that is, in order to cause Smith to wince. The instantiation (or potential instantiation) of the core realizer for hunger need not form any part of the background conditions for the neurological causes and effects of \textit{c-fiber firing} in order for the core realizer of hunger to form part of the surround that allows \textit{c-fiber firing} to play the causal role of pain.
Now we are in a position to establish the first thesis of this paper: surrounds are not causal background conditions. This I will do in the next section.

§4

In order to show that a core’s surround is not the same thing as its causal background conditions it will be helpful to have before us the diagram in figure 1.

In Figure 1, the four small circles represent core realizers: neurological properties had by some organism. For example, the circle labelled “Pj” represents the organism’s core realizer for pain. The other small circles represent core realizers of other psychological states. The large circle that encompasses the four small circles represents a surround. The solid arrows represent a causal chain on a given occasion of a realization of pain. We may imagine the situation depicted in Figure 1 like this: the solid inbound arrow indicates skin damage, which causes an instance of
P_j, the core realizer for pain. The instance of P_j, in turn, has some immediate external or behavioral effect (the upper outbound arrow), as well as an internal effect: P_j causes P_m. P_m in turn causes P_n, which in turn has an external or behavioral effect of its own (the lower outbound arrow).

Now, it forms no part of the causal explanation of the outbound arrows in Figure 1 that the entity is capable of having P_k, and that instances of P_k, when they occur, both plays the causal role of some mental state and tend to cause instances of P_n. These facts about P_k form no part of the background conditions for the occurrence of the lower outbound arrow depicted in Figure 1. Nevertheless, these facts about P_k are part of the surround that ensures that the core realizers P_j, P_m, and P_n play the psychological causal roles that they play. The facts about P_k are part of what makes P_n a core realizer for some psychological property and these facts about P_k are also implicated in P_j’s being a core realizer for pain, even though P_k’s potential arrows never touch P_j directly, nor vice versa. This is because P_k is part of the n-tuple of core realizers that satisfies the Ramsey-Lewis sentence of the overall psychological theory. (Less abstractly, it is because P_k’s potential arrows help to create the role for P_n, and P_n’s having the role it does is integral to P_j’s having the role it does.) P_k is part of P_j’s surround and therefore part of the total realizer.

The key point is that P_k has nothing to do with underwriting the causal powers of P_j. P_k is no part of the background conditions for the causal powers of P_j -- yet P_k is part of the explanation for why P_j plays the causal role of pain. P_k -- or rather the capacity to have P_k under appropriate conditions -- is part of the surround depicted in Figure 1; it is part of the determinate form of the property: being physically constituted in such a way that P_j plays the causal role definitive of pain. So even on the occasion depicted in Figure 1 the facts about P_k are relevant to
explaining why an instance of pain occurs, even though the facts about P_k form no part of the explanation for why an instance of P_n occurs.

This is why surrounds are not background conditions: a surround will never form part of the background conditions for any occasion-specific neurological cause or effect. A surround is not an INUS condition: not an insufficient but necessary part of a sufficient but unnecessary condition for a causal effect. Surrounds are necessary for the existence or individuation of realized properties, not for their effects. They are necessary for causal role-playings, not for causings.

§5

The upshot of the discussion in §4 is that a core and its surround are not the same thing as a core and its causal background conditions, and that means in turn that total realizers -- the conjunctions of cores and their surrounds -- are not causally relevant to a core’s effects. *Thus the properties that strongly subvene realized properties are not the same as the properties that provide causal explanations of a realized properties’ effects. Indeed, the properties that strongly subvene realized properties don’t have any causal effects at all.*

This is in contradiction to much of the work that has been done by reductivists and non-reductivists in the philosophy of mind for many years. To give a particularly straightforward example, it is in contradiction to Jaegwon Kim’s Causal Inheritance Principle -- the principle that drives his well-known Disjunction Argument.

[The Causal Inheritance Principle] If mental property M is realized in a system at t in virtue of physical realization base P, the causal powers of *this instance of M* are identical with the causal powers of P. (Kim, 1992: 326; italics original)
Many of Kim’s non-reductivist opponents accept this principle. But it cannot be right. As has been shown by others, by “realization base $P$” Kim has to mean $M$’s total, not core, realizer. But I have shown that surrounds do not figure in the causal explanation of the effects of the instantiations of core realizers in functional systems -- does not figure in the explanation of the arrows in Figure 1, for example. And if surrounds are not part of the explanation of what causal proceeds from the instantiation of sufficient conditions for an instance of a realized property, then total realizers are not part of that explanation, either.

What total realizers do explain are causal role-playings. But causal role-playings need not entail and do not explain any causings. Here is another way of making this specific point. An instance of pain will occur in an organism when that organism both satisfies $T$ -- that is, when it has a surround -- and is caused to have its core realizer for pain. The cause of the core realizer need not be standard -- the cause need not be a cause that is part of pain’s “most typical syndrome of causes and effects.” Electrode stimulation will do. All that is needed in order for pain to be realized is for the organism to have $T$ and for an instantiation of the core realizer to be caused. Furthermore the instantiation of the core realizer need not have any standard, role-relevant effect, or even effect at all. It doesn’t matter -- at the moment of the core realizer’s instantiation its property instance is playing the causal role of pain. For a property to play the causal role of pain is simply for it to be had by an organism for which it is the pain-th member of a $T$-satisfying n-tuple. It doesn’t matter what effects it has. Further, any effect it does have may be explained as effects of the core realizer plus the background conditions that obtained on the occasion of the

---

4 See Antony and Levine (1997); Witmer (2003). The Disjunction Argument only gets off the ground if we assume that the disjunction in question is the disjunction of a realized property’s total realizers, not the disjunction of its core realizers.
core realizer’s instantiation. For reasons that were explained in the previous section, the fact that a given $T$-satisfying n-tuple is had by an organism is never part of the background conditions that explain a core realizer’s effects. That means, ultimately, that the conjunctions of cores and surrounds don’t make anything happen causally. That is to say, total realizers don’t make anything happen causally; they figure in realizings, not in causings. This establishes (T2).

§6

As I mentioned previously, T1 and T2, especially T2, are what I take to be the philosophically most interesting and consequential results we can get from a correct understanding of total realizers. Nevertheless I think it is worth the time to draw out a couple of further consequences. Seeing these will help us to understand what total realizers are and how they work within the causal role functionalist system.

Misunderstandings of total realizers are very common. Even if they are not instantly consequential within the context of the paper in which they are being presented, such misunderstandings may indicate trouble down the road. Thus even a philosopher as careful as Gene Witmer may make an error that, given our discussion, is pretty straightforward.

If $N_1$, $N_2$, and so on are all the physical properties capable of core realizing pain, then the list of total physical realizers ($R_1$, $R_2$, . . .) can be expressed thus:

- Having $N_1$ and being such that one’s having $N_1$ meets condition $\psi$.
- Having $N_2$ and being such that one’s having $N_2$ meets condition $\psi$. 
Here, Witmer claims that a total realizer is a conjunctive property whose conjuncts are a core realizer and the property of being such that that core realizer plays the relevant causal role. This is a very common way to picture total realizers, but it is mistaken. It entails that for any realized property there are exactly as many total realizers of the realized property as there are core realizers, which is false, and it entails a denial of T3, the thesis that two organisms may have the same core realizer for some mental state M but different surrounds and therefore different total realizers for M. The key point is that Witmer asserts a claim that entails both that any given realized property will have the same number of total as core realizers and the denial of T3 because Witmer takes surrounds to be less determinate properties than they are.

A surround is specific n-tuple of properties. It is not the property of having a specific core for a given role plus any other group of core realizers that complete the n-tuple. That, though, is often what philosophers take surrounds to be. In this case, Witmer cites a 1997 article by Louise Antony and Joseph Levine as the source of his understanding of surrounds; two of the most prominent philosophers in the discussion of causal role functionalism (Witmer, 2003: 66). Yet their understanding of surrounds and therefore of total realizers is informed by a simple misreading of Shoemaker’s 1981 article. Antony and Levine write:

Shoemaker suggests that we think of Tx as expressing the property “being physically constituted in such a way that Pj plays the causal role definitive of pain.” (Shoemaker, 1981: 265) This conjunctive property [viz., (Tx & Pj)] constitutes what Shoemaker calls the

---

5 Here, the phrase “being such that one’s having N1 meets condition ψ” means “being such that one’s N1-ness plays the appropriate causal role.” The condition is a causally characterized condition; a causal role.
“total realization of the second-order, functionally defined property. (Antony and Levine, 1997: 89)\(^6\)

This is a mistake. Antony and Levine have simply left off a key part of Shoemaker’s sentence. Here is what Shoemaker wrote:

Whether A is in pain depends on whether he has \(P_j\), and clearly his having \(P_j\) is in some sense constitutive of his being in pain, as opposed to being a cause of it. But this is true only because A has the property expressed by ‘\(T_x\)’; this can be thought of as a **determinate form of the property**: being physically constituted in such a way that \(P_j\) plays the causal role definitive of pain. (Shoemaker, 1981: 97; Shoemaker’s italics; my boldface)

Once we correct for Antony and Levine’s error, we must make an adjustment to Witmer’s passage on the relation between core and total realizers. Evidently there will be more than one determinate form of the property being physically constituted in such a way that \(P_j\) plays the causal role definitive of pain. There will be as many ways as there are possible ordered n-tuples of properties that both satisfy \(T\) and have \(P\) as the \(\text{pain}\_\text{th}\) member. But that means that it is not the case that each core realizer for pain corresponds in a one-to-one way with a single total realizer for pain.

I point this out because it means that on the canonical understanding of surrounds and total realizers, the one that everyone cites, including Shoemaker himself, it is entirely possible for two organisms to have the same core realizer for a given mental property but different total realizers for that mental property.\(^7\) Understanding this may not be urgent to any particular philosophical project that makes use of core and total realizers, but understanding it does seem to

---

\(^6\) Antony and Levine’s citation of Shoemaker is a little hard to interpret. The year for their citation corresponds to the original publication of Shoemaker’s “Some Varieties of Functionalism” in Philosophical Topics (Shoemaker, 1981) but the page number that they cite corresponds to the reprint in Shoemaker’s anthology Identity, Cause, and Mind (Shoemaker, 2003). It doesn’t matter; the same wording about determinateness occurs in both.

\(^7\) Shoemaker refers us back to “Some Varieties of Functionalism” in his discussion of surrounds and total realizers in the recent Physical realization (Shoemaker, 2007: 21, f. 14).
be requisite for a complete understanding of how the theory was supposed by its founders to work. The canonical text on total realizers supports T3.

It should be clear, though, that whether one keeps or discards Shoemaker’s take on the determinateness of surrounds has no bearing on the correctness of T1 or T2. If one drops the clause “a determinate form of the property” from the description of surrounds, as do Antony and Levine, one will have, if anything, even less reason to suspect that the surround forms part of the background conditions for the causal efficacy of the core realizer on any given occasion of the core’s instantiation.

§7

In this section I will discuss T4, the thesis that all of a given entity’s core realizers have the same surround. I take it that this is nearly evident at this point. A surround is a property of having a specific n-tuple of entities at the ready to be had (to “fire” to borrow the language of c-fibers) in causal patterns described by T. One and the same such property accompanies each instantiation of each core realizer to form a total realizer in a given entity.

Take for example Smith, a typical human who has just mildly stubbed her toe. Smith is disposed to have the n-tuple of core realizers that is typical for humans and therefore she has the property of having that specific n-tuple of core realizers at the ready. When she stubs her toe this causes her to have the property of c-fiber firing which in turn causes her to have the property of, let us say, m-fiber firing, the core realizer for humans of the belief that one is pain. So in Smith’s ordered n-tuple, c-fiber firing is the painth member, and m-fiber firing is the beliefthatpainth member, as it were. It is the property of having this same n-tuple that conjoins with, now c-fiber
firing, now m-fiber firing, to form the total realizers of pain and the pain-belief, respectively. There are two reasons, I believe, to find this implausible, both of which are mistaken.

First, one might think that for each core realizer, the relevant ordered n-tuple is not the complete n-tuple of properties that satisfies \( T \) but that n-tuple minus the core in question. To put that more precisely, since the surround is the property of having an n-tuple of properties “at the ready” to be instantiated in \( T \)-appropriate patterns, and since at the moment Smith stubs her toe (to revert to our example) the property of c-fiber firing is not “at the ready” but actually in action, it would seem to be inappropriate to count Smith as having any property that includes the property of having c-fiber firing “at the ready.” If that is right then each total realizer for Smith will include a core realizer and a surround that includes all of Smith’s core realizer except the one in the first conjunct. But I think that this is the wrong way to understand core and surrounds.

I have been using the phrase “at the ready” in my description of surrounds because it gets the point across that it is not necessary to have every single member of a \( T \)-satisfying n-tuple actually instantiated at the same time in order to have the surround at that time. Indeed, it is not necessary to have any member of a \( T \)-satisfying n-tuple at a given time in order to have the surround that is characterized by that n-tuple at that time. What is needed is to be such that the properties that make up the n-tuple are the ones that would tend to be had in response to \( T \)-relevant circumstances. And it remains true of Smith that she is such that she tends to have c-fiber firing in response to pain appropriate circumstances even at the moment that she is having c-fiber firing in response to a pain-appropriate circumstance.

The second reason to find T4 implausible is to take the wording of the canonical description of surrounds too literally. Now, I just got done in the previous section, chastising
other philosophers for not taking the wording literally enough. So I had better, here, have a good reason to be making a suggestion that goes in the opposite direction. I believe that I do. Here again is the canonical wording:

\[ Tx \] can be thought of as a determinate form of the property: being physically constituted in such a way that \( P_j \) plays the causal role definitive of pain. (Shoemaker, 1981: 97; italics original)

It might be thought that Smith’s surround for pain, \( Tx \), is specific to pain because the word “pain” occurs in the description Shoemaker gives. Thus, the thought would be that these two properties of Smith are not identical:

- **\( T_{x_{Smith1}} \)** Smith’s determinate form of the property: being physically constituted in such a way that \( P_j \) plays the causal role definitive of pain.
- **\( T_{x_{Smith2}} \)** Smith’s determinate form of the property: being physically constituted in such a way that \( P_k \) plays the causal role definitive of hunger.

But they are identical. The two definitions are just different ways of picking out the same property, \( Tx \). If \( Tx \) were not determinate to at least the grain of the sentence the spells out \( T \) then \( T_{x_{Smith1}} \) and \( T_{x_{Smith2}} \) could come apart. To see this, simply drop the words “determinate form of the” from each definition.

- **\( T_{x_{Smith1}^*} \)** Smith’s property: being physically constituted in such a way that \( P_j \) plays the causal role definitive of pain.
- **\( T_{x_{Smith2}^*} \)** Smith’s property: being physically constituted in such a way that \( P_k \) plays the causal role definitive of hunger.

It is of no accord to Smith’s having \( T_{x_{Smith1}^*} \) that \( P_k \) specifically plays the role of hunger, in Smith. All that matters to Smith’s having \( T_{x_{Smith1}^*} \) is that Smith has some property or other -- or, more precisely, that Smith is capable of having some property or other -- that acts
as the core realizer of hunger. Similarly, it is of no accord to Smith’s having $T_x Smith_2^*$ that $P_j$ specifically plays the role of pain, in Smith. All that matters to Smith’s having $T_x Smith_2^*$ is that Smith has some property or other -- or, more precisely, that Smith is capable of having some property or other -- that acts as the core realizer of pain.

Contrariwise, Smith would not have $T_x Smith_2$ were Smith’s property that stood ready to play the role of hunger not $P_k$, and Smith would not have $T_x Smith_1$ were Smith’s property that stood ready to play the role of pain not $P_k$. Thus, T4 is born out by the canonical account of surrounds and total realizers.

§8

It appears to be true that T3 and T4 stand or fall together. Whether they stand or fall depends upon whether one takes surrounds to be as determinate as Shoemaker took them to be in his originating 1981 article, or less determinate. If one takes surrounds to be as determinate as did Shoemaker, then T3 and T4 stand. If one takes surrounds to be less determinate, then they fall.

I want to stress, though, that T1 and T2 stand regardless. Less determinate forms of surrounds are not background conditions for the causal explanations of the occurrences of core realizers or of the effects of the occurrences of core realizers any more than are more determinate forms of surrounds. But this means that the conjunction of a core realizer and a surround is not the same thing as and has nothing much to do with the conjunction of a core realizer and the background conditions that fix its efficacy on any given occasion. But another name for the
conjunction of a core realizer and its surround is *total* realizer. Taking all of this together, we can see that the following principle is just false:

*The Principle of Causal Subveners* For any instance of a role property R, the property P upon which the instance of R strongly supervenes, its total realizer, figures in the causal explanation of the effects of that instance of R.

The falsity of this principle -- a principle so ingrained in the literature that it is never stated explicitly -- is disastrous for causal role functionalism. Debates within causal role functionalism presuppose that the property that strongly subvenes a realized property is the same property as the property that plays its causal role. Causal role functionalists will grant that there must be two senses of “playing a causal role” since it must be allowed that both the core and the total realizer of a realized property play its causal role in some sense. But it is a minimal condition to being any sort of causal role player that a property figure in the causal explanation of the effects of a realized property. But as we can see there is no sense at all in which total realizers meet that condition.
Chapter Three
Kim’ Supervenience Argument and the Nature of Total Realizers

§1

The problem of causal exclusion has become something of a perennial thorn in the side of philosophers of mind who wish to hold that mental states are causally efficacious yet irreducible to physical states. The problem, as is well known, is that once all of the world’s physical states, properties, events, etc., have completed their causal work there seems to be nothing in a physicalistic world for unreduced mental states, properties, events, etc., to do. We are told by advocates of the exclusion problem that because every event in a physicalistic world has a physical cause, any event that someone cares to claim has been caused by the mental goings-on in a physicalistic world will have already been caused by the physical goings-on in that world. Furthermore, it seems objectionable to try to get around this problem by claiming that some states, properties, events, etc., are caused twice-over, with mental causation being responsible for the superfluous “push.” Given all of this, what’s a mind to do in a physicalistic world? Worse, since it seems likely that our world is a physicalistic world, what’s a mind to do in our world?

The most trenchant presentation of the causal exclusion problem over the past decade or so has been Jaegwon Kim’s. He presents the exclusion problem via the by-now familiar “Supervenience Argument”; offered most notably in his books Mind in a Physical World (Kim, 1998) and Physicalism or Something Near Enough (Kim, 2005). Kim uses the Supervenience Argument to show that non-reductivism about mental states is untenable. He thereby hopes to

---

1 A paper closely based on this chapter is forthcoming from the European Journal of Philosophy.
2 This “already” is an “already” of priority, not temporality.
motivate the adoption of a reductive view of mind. In response, non-reductivists have tried in several ways to refute or problematize the Supervenience Argument.

In this chapter I present what will look at first like yet another attempt at refuting Kim’s Supervenience Argument on behalf of the non-reductive physicalist. But my goal is not so sanguine. The issue I discuss will be a problem both for Kim’s Supervenience Argument and for the non-reductivist view that Kim’s Supervenience Argument is meant to attack. I will argue that the reductivists and the non-reductivists in this ongoing debate over the exclusion problem share a flawed assumption about the nature of the base physical properties upon which mental properties supervene. Once this flawed assumption is exposed, it is none too clear that either side’s view can be sustained in the form that has heretofore been common.

I will proceed as follows. In §2 I discuss the general form of the exclusion problem. In §3 I review Kim’s Supervenience Argument, drawing attention to a hidden assumption; namely, that the supervenience of mental properties on physical properties is to be explained by the realization of mental properties by physical properties -- where realization is explicated as causal role-playing. In §4 I present a novel response to the Supervenience Argument that exploits this hidden assumption. I argue that an incoherence is buried in the way the Supervenience Argument asks us to conceive of the base physical properties upon which mental properties supervene; we are asked to think of the base properties both as causal role-players -- and therefore as typical physical/causal properties -- but also as properties that are unconditionally sufficient for the mental properties that they subvene. These requirements are mutually exclusive. Therefore the Supervenience Argument fails. I will close in §6 with a discussion of
why the argument presented in §5 has disruptive (though maybe fruitfully disruptive) consequences for both the reductivist and non-reductivist sides of the debate.

§2

The problem of causal exclusion can be put more formally than was done in §1 by making a list of claims about mental causation and the nature of the physical world that are individually agreeable but nonetheless inconsistent with one another. Several philosophers have formalized the exclusion problem by making lists of this sort.³ Let us make use of Karen Bennett’s list, from her paper “Exclusion again”:

<table>
<thead>
<tr>
<th>Distinctness:</th>
<th>Mental properties (and perhaps events) are distinct from physical properties (or events).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness:</td>
<td>Everything that happens has a sufficient physical cause.</td>
</tr>
<tr>
<td>Efficacy:</td>
<td>Mental events sometimes cause physical ones, and sometimes do so in virtue of their mental properties.</td>
</tr>
<tr>
<td>Nonoverdetermination:</td>
<td>The effects of mental causes are not systematically overdetermined; they are not on a par with the deaths of firing squad victims.</td>
</tr>
<tr>
<td>Exclusion:</td>
<td>No effect has more than one sufficient cause unless it is overdetermined. (Bennett, 2008: 280)</td>
</tr>
</tbody>
</table>

The problem of causal exclusion, then, is that at least one of these five claims has to go.

Here is a proof. Let us assume that every cause is part of a sufficient cause. Then

Nonoverdetermination and Exclusion together guarantee that mental causes of events (if there are any) are parts of unique sufficient causes of those events. Completeness says that every event has a sufficient physical cause; together with Nonoverdetermination and Exclusion this means that mental causes (if there are any) are parts of unique sufficient physical causes. But

Efficacy states that there are mental causes; so mental causes exist and are parts of physical causes. Assume that every part of a physical cause is physical. It follows that mental causes are physical. But this contradicts Distinctness.

So it seems that if we want to have so much as a shot at a coherent metaphysics of mind we will have to drop one of the five claims; which one we choose to drop will determine what our general view of mind is. If we drop Distinctness we become reductionists. If we drop Completeness we seem to court some sort of interactionist dualism. If we drop Efficacy we become epiphenomenalists. If we drop either Nonoverdetermination or Exclusion we become compatibilists -- philosophers who think that we can both physically have our mental cake and non-reductively eat it too. Compatiblism is the most common sort of move made by non-reductive physicalists in response to the exclusion problem. Bennett points out that the exclusion problem was introduced into contemporary philosophy by Norman Malcolm in his “The Conceivability of Mechanism” (Malcolm, 1968) and that a compatibilist response was attempted right away by Alvin Goldman in “The Compatibility of Mechanism and Purpose” (Goldman, 1969) (Bennett, 2003: 483). Bennett does not add, perhaps because she assumes it is well-known, that Malcolm did not use the exclusion problem to motivate physicalism, as would later of its advocates such as Kim; rather, Malcolm introduced the exclusion problem in order to deny Completeness (to put the matter in terms of Bennett’s above list). Goldman offered the denial of Nonoverdetermination as an alternative to the denial of Completeness. The significant difference between Malcolm’s and Goldman’s solutions is that Malcolm took his denial of Completeness to be a denial of physicalism (i.e. a

---

4 In addition to Bennett herself (Bennett, 2003, 2008), notable recent compatibilists include Block (Block, 1990, Block, 1997), Noordof (Noordhof, 1997), and Antony (Antony, 2003).
denial of “The Conceivability of Mechanism”) while Goldman wished to keep physicalism (or at least its conceivability) while preserving the causal relevance of the mental. This will be of some interest later on as we will see in §6 that I, like Malcolm, am going to deny Completeness. My denial, however, will not contravene the correctness or conceivability of physicalism.

§3

In this section I will briefly lay out Kim’s Supervenience Argument. In particular, I wish to draw attention to a hidden assumption in the Supervenience Argument that will be important to the critique I will present in §4. But before we get the hidden assumption let us remind ourselves of the Supervenience Argument as a whole; the idea behind the argument is captured by the familiar diagram in Figure 1.

![Figure 1](image)

M and M* are mental properties of some subject, say Smith. P and P* are physical properties of Smith. The vertical lines indicate supervenience -- more specifically, strong supervenience: M strongly supervenes on P and M* strongly supervenes on P*. The solid horizontal arrow running from P to P* is a causal arrow; it indicates that the physical property P of Smith causes Smith to have the property P*. Finally, the dashed arrow running from M to M*
is the point of the argument; its status as a genuine causal arrow is what the Supervenience Argument assumes in order to deny.

Here are the steps in the argument, from Kim’s 2005 presentation in *Physicalism or Something Near Enough*.

(1) M causes M*.
(2) For some physical property P*, M* has P* as its supervenience base.
(3) M caused M* by causing its supervenience base P*.
(4) M has a physical supervenience base, P.
(5) M causes P*, and P causes P*.
(6) M \neq P.
(7) P* is not causally overdetermined by M and P.
(8) The putative mental cause, M, is excluded by the physical cause, P.

That is, P, not M, is a cause of P*.\(^5\)

The argument works by showing that the assumption that M causes M* gets the result that M’s supervenience base, P, causes M*’s supervenience base, P*. Since, by definition of supervenience, P* guarantees the instantiation of M*, and since P* can’t plausibly have more than one cause, it follows that -- contrary to our original assumption -- M does not really have anything to do with bringing about M*.\(^6\)

Philosophers have tried several responses to the Supervenience Argument over the years. By far the most common response is to challenge the claim that P* can’t plausibly have more than one cause. This, of course, is an expression of the compatibilist response to the Exclusion

---

\(^5\) All steps are quoted from (Kim, 2005: 39-43). Italics are from the original.

\(^6\) Here is the definition of supervenience that Kim employs in the 2005 presentation of the argument:

*Supervenience*. Mental properties strongly supervene on physical/biological properties. That is, if any system \(s\) instantiates a mental property \(M\) at \(t\), there necessarily exists a physical property \(P\) such that \(s\) instantiates \(P\) at \(t\), and necessarily anything instantiating \(P\) at any time instantiates \(M\) at that time. (Kim, 2005: 33)

The definitions of supervenience employed in other presentations of the Supervenience Argument, most importantly in the 1998 *Mind in a Physical World*, are equivalent. Note that the kind of supervenience in question is strong supervenience.
Problem that we saw in §2; there, compatiblism appeared as a denial of either Nonoverdetermination or Exclusion; here, it appears as a denial of (7) or (8). Other common responses to the Supervenience argument are the generalization response (the idea that the Supervenience Argument can’t be right because if it is right is generalizes to all special science properties, an allegedly intolerable result) and what I’ll call the programmatic response (the idea that the Supervenience Argument relies upon too restrictive a notion of causation). 7

In the next section I will present a different kind of response to the Supervenience Argument. To prepare the way, I wish here to attend to the move from step (3) to step (5); from the claim that in order for M to cause M* M must cause M*’s supervenience base, P*, to the claim that M’s supervenience base, P, causes P*. Kim writes that we should accept that P causes P* because of something like the transitivity of nomological sufficiency (2005: 41): since P is nomologically sufficient for M (by step 4) and M is nomologically sufficient for P* (by step 3), therefore P is nomologically sufficient for P*. Let me set this brief argument out to highlight it.

\[
\begin{align*}
(CS\ 1) & \quad P \text{ is nomologically sufficient for } M. \\
(CS\ 2) & \quad M \text{ is nomologically sufficient for } P*. \\
(CS\ 3) & \quad \therefore \ P \text{ is nomologically sufficient for } P*. 8
\end{align*}
\]

Over the next few paragraphs I will first present a reason to think that this mini-argument (CS 1-3) is wrongheaded. I will then then show how to save it, but at the cost of opening the Supervenience Argument as a whole to the response I will present in §4.

---

7 For the generalization argument, see Noordof (1999), Gillett and Rives (2001), and Bontly (2002). For the programmatic response see Walter (2008: 681) and Loewer (2002).

8 Kim makes the same argument in the earlier *Mind in a Physical World:*

If you take causation to be grounded in nomological sufficiency, P qualifies as a cause of P*, for, since P is sufficient for M and M is sufficient for P*, P is sufficient for P*. (Kim, 1998: 43; italicization of "M" and "P" removed for consistency)
First, the problem with (CS 1-3). On ordinary non-reductive-physicalist assumptions about the relationship between mental and physical properties, M is not nomologically sufficient for P*, so (CS 2) is false and (CS 1-3) is unsound. M is not nomologically sufficient for P* because, given that mental property M* is multiply realizable, M* will have many possible strong supervenience bases. The fact that mental property M is nomologically sufficient for M* (by the standard functionalist construal of M) is not enough to show that M is sufficient for any specific one of M*’s strong supervenience bases. M’s sufficiency for M* entails M’s sufficiency for some-one-or-another of M*’s strong supervenience-bases but it does not entail sufficiency for P* in particular (for example). We have no reason to think that M causes P*; therefore, we have no reason to think that P causes P*. So (CS 3) and therefore step (5) in the Supervenience Argument are both false.

The objection I just raised may seem to miss the point; after all, step (3) of the Supervenience Argument is the assertion that M causes P*, so there should be no problem, now, in drawing upon that assertion for (CS 2) and the line of reasoning that results in (5). But my point is that (3) is precisely the claim that we should find dubious, given M*’s multiple realizability. What is the justification for (3) supposed to be? In Physicalism or Something Near Enough Kim states that M’s causing P* is the “simple and natural” way to assure that M’s causal powers are relevant to the instantiation of M* (2005: 40). Kim’s argument is that given that P* is what strongly subvenes M* on this occasion, it had better be the case that on this occasion M brings about P*’s instantiation. But again, this is what seems out-of-bounds, given M*’s multiple realizability.
The solution to this -- the way, temporarily, to save the Supervenience Argument -- is to take care in distinguishing between properties and property instances. While it is true that property M is not in general sufficient for any particular one of M*’s strong supervenience bases, and in particular is not sufficient for P*, Kim’s claim is just that this instance of M is sufficient for P*. Kim’s slightly hidden reasoning, I believe, is that on this occasion the M-instance inherits the causal powers of its supervenience base, P. This is what justifies the claim in step (3) that M causes P*. But now we get to the rub.

The Supervenience Argument is supposed to show that the causal powers of unreduced mental properties are excluded or preempted by the causal powers of their physical bases; it is supposed to show this by assuming only the principle of Supervenience. But -- and this is the key point -- by importing the idea that the M-instance inherits its causal powers from its supervenience base P, Kim is assuming an explanation for why M strongly supervenes on P. M strongly supervenes on P because P realizes M, where “realizes” means “plays the causal role of.” This is a much stronger assumption than the minimal physicalist assumption that the mental strongly supervenes on the physical -- it is a move from mere strong supervenience to the realization relation and, indeed, to a particular explication of realization as causal role-playing. This is the idea that I will exploit in the next section.

---

9 Another way of making this point: the Supervenience Argument is not supposed to rely upon Kim’s familiar Causal Inheritance Principle. But, actually, (CS 1-3) does rely upon it:

[The Causal Inheritance Principle] If mental property M is realized in a system at t in virtue of physical realization base P, the causal powers of this instance of M are identical to the causal powers of P. (Kim, 1993a: 326)

Once we adopt the Causal Inheritance Principle for the property M in Figure 1, there is no mystery about why M should cause P* in particular, out of all of M*’s possible strong supervenience bases: M does because P does.
§4

In §3 I argued that the Supervenience Argument makes a covert appeal to the realization relation as the relation that explains the supervenience of mental properties on physical properties. Without that assumption there is no way to arrive at P’s causing P* from the stipulation that M causes M*. It is time to put this line of reasoning to work. Briefly, the difficulty that the Supervenience Argument faces is that it assumes both that P and P* as depicted in Figure 1 are strong supervenience bases for M and M* and (covertly) that P and P* play the causal roles of M and M*. The problem with this dual assumption is that the properties that play the causal roles of M and M* (i.e. M and M*’s core realizers) are not unconditionally sufficient for M and M*. And the properties that are unconditionally sufficient for M and M* (their total realizers) do not play their causal roles. The upshot of this is that there do not exist any Ps of the sort that the Supervenience Argument needs in order to go through.

In order to make the case for the line of thought I just laid out I will first spend a few paragraphs discussing the distinction between core and total realizers and how our understanding of them should affect the our understanding of Kim’s argument.

The distinction between core and total realizers was first drawn by Sydney Shoemaker in his important 1981 article “Some Varieties of Functionalism” (Shoemaker, 1981). Shoemaker defined core realizers using David Lewis’s (1983b) method of specifying properties via their occupancy of causal roles, a method which in turn uses the idea of a Ramsified theory to say what the causal roles are. A core realizer of, say, a psychological property such as pain, is a
member of an n-tuple of physical properties had by a subject, say Smith, that causally interact in such a way that the members of they, the members of the n-tuple, together satisfy the psychological theory T. Smith can’t have pain without having the whole psychological theory satisfied, and that means that she can’t have pain unless she has or anyway is capable of having all of the members of an appropriate n-tuple of properties. Each of those members, again, is a core realizer of some psychological property.

To say that some physical property is a member of Smith’s n-tuple of core realizers is to say that in Smith that physical property is caused by and causes other members of the n-tuple in ways that obey patterns described by T. Taking again the standard example of c-fiber firing and pain, to say that c-fiber firing is the core realizer of pain in Smith is to say that c-fiber firing is the \(j\)th member of Smith’s n-tuple of physical T-satisfying properties, where the \(j\)th member is the property that is caused by (inter alia) skin damage and that causes (let us say, inter alia) the \(k\)th member of the n-tuple, where the \(k\)th member is the one that plays the causal role of the belief that one is in pain.

This is not to say that instances of c-fiber firing are unconditionally sufficient for instances of pain. The mere fact that Smith instantiates c-fiber-firing does not guarantee that she instantiates pain. To get the property that is unconditionally sufficient for pain -- to get property such that if Smith had it it she would for sure have pain -- we have to find the conjunctive property that is the combination of c-fiber firing together with all of the properties of Smith and possibly the world that guarantee that instances of c-fiber firing in Smith play the causal role of pain. What are those additional properties?
Since the theory T defines psychological properties holistically, we first of all we need to add the whole n-tuple of Smith’s properties that satisfy T. Or, rather, we need Smith’s property of having this n-tuple in appropriate, T-obeying patterns. In Shoemaker’s 1981 article on core and total realizers he called this (very complex) additional property “T(x)” and the total realizer a conjunctive property whose conjuncts are, at least, the core realizer and T(x) (Shoemaker, 1981: 96-97).10

What else do we need to add to Smith’s core realizers in order to get properties that are unconditionally sufficient for Smith’s realized properties? In a later book Shoemaker wrote about properties that we may need:

Assuming externalism about content, having thoughts about water will require (at least to a first approximation) living in an environment in which there is water. And in that case the total realizer for the thought that there is water in the glass will be partly a relational property that something has only in an environment in which there is, or has been, water. If it is a requirement for having such a thought that the subject should in the past have interacted in certain ways with her environment, or with other members of her linguistic community, then the total realizer will be not only in part a relational property but in part an historical property. (Shoemaker, 2007: 22)

Shoemaker writes as if for some mental properties but not for others we must incorporate properties of the environment in order to get the mental property’s complete total realizer. For example, Shoemaker writes as if for the belief that there is water in the glass we need a “surround” (the “surround” part of a total realizer is everything but its core) that extends into the environment, and perhaps into the past, but for a “property like being in pain it seems plausible

---

10 A clarificatory point about T(x): Smith does not need to have all of her core realizers active (i.e. instantiated) at once in order to have either T(x) or the total realizer for any one mental property -- she just has to be able to have the n-tuple of core realizers, and to be able to have them in T-obeying patterns. A full account of how to define T(x) is beyond the scope of this paper, but I hope this is enough to give the general idea.
that that the ‘surround’ part of the total realizer will be internal to the subject” (Shoemaker, 2007: 22). But it is not clear that a functionalist may be this piecemeal about realization. Shoemaker acknowledges, “to the extent to which the domain of mental properties is holistic, the surrounds of total realizers of mental properties will overlap” (2007: 22). Indeed.

Consider the property of being thirsty. Thirst might at first glance seem like as good a candidate as pain for being a mental property whose total realizer is wholly “internal to the subject.” But if we think that *causes a desire for water* it is part of the causal role of the property of being thirsty -- part of thirst’s “syndrome of most typical causes and effects” to use the language of Lewis’s originating paper on this topic (Lewis, 1983a: 100) -- then we should think that what Shoemaker calls “the surround part of the total realizer” for thirst extends into the environment, after all. One can’t have thirst if one can’t desire water, and one can’t desire water unless one’s environment (currently or historically) has water.

Leaving the specific case of thirst aside, note more generally that since the property of having a complete n-tuple of core realizers that causally interact in T-appropriate ways was, recall, the *first* thing we understood that we needed to add to any given core realizer to get a total realizer, we see that the functional conception of mental properties is actually *completely* holistic. This means that all of Smith’s core realizers must in fact have the same surround. If the total realizers of one mental state extends into the environment and into the past, then they all do.

But we are not done yet. Is there anything *else* that we must add to Smith’s core realizers to make them unconditionally sufficient for her mental properties? That depends upon what our theory of properties is. If we hold that a property has the causal powers that it has essentially, in every possible world in which the property exists at all, then we probably don’t need to add
anything else to guarantee that c-fiber firing plays the causal role of pain. But if we think that what c-fiber firing does in a world depends on what the laws of nature are in that world, then we must add the holding of appropriate causal laws to the batch of things that we must combine with c-fiber firing to guarantee an occurrence of pain. That is, Smith’s property of being such that the actual laws of nature hold must be added to her total realizer for pain.

Shoemaker acknowledges the point about laws of nature early on in his recent book on realization (2007: 6); he intends it merely as a recommendation for his own theory of properties, according to which properties (such as core realizers) have their causal powers essentially (thus obviating the need to add the holding of laws of nature to the total realizer). But Brian McLaughlin points out in a review of Shoemaker’s book that the issue is rather more serious. McLaughlin writes that if we don’t adopt Shoemaker’s full-blown theory of properties (as most philosophers are not willing to do) then we must add the property of being such that the actual laws of nature hold to the total realizer. Once we do this, the resulting total realizer is not a causal property -- is not the kind of property that can cause or be caused by anything:

Suppose CTP [Shoemaker’s causal theory of properties, according to which properties have their causal powers essentially] is false. Then, we can hold, Shoemaker says, that included in the total realizer of a property P are the obtaining of causal laws that determines P’s causal profile. Label the conjunction of the laws in question, L. Although Q can realize P without instances of Q metaphysically necessitating instances of P, the total realizer of P will be being Q and such that L. Further, the instantiation of the property of Q and being such that L will metaphysically necessitate

11 Shoemaker writes:

It should be noted that unless one holds (as I do, but will not insist on here) that properties have their causal profiles essentially, one cannot hold that the instantiation of a realizer property is, in and of itself, sufficient for the instantiation of the property it realizes . . . We can get around this by including in the realizer the obtaining of a set of causal laws -- normally the laws that obtain in the actual world. (Shoemaker, 2007: 6)
an instance of $P \ldots$ The total realizer is thus a conjunctive property that includes certain causal laws as conjuncts. This proposal faces a problem. Presumably, a total realizer is a realizer. To be a realizer, however, a property has to have a causal profile. How could being $Q$ and such that $L$ have a causal profile? Causal laws themselves (even if they are singular relations among properties) are never causes or effects. (McLaughlin, 2009)

The principle that McLaughlin adopts in order to deny the causal efficacy of non-Shoemakerian total realizers seems to be this:

\[(CCP) \quad \text{A conjunctive property has causal powers if and only if each of its conjunct properties has causal powers.}\]

Once we adopt CCP, McLaughlin’s line of reasoning follows easily. Since the property (call it “$L$”) of being such that the actual laws of nature hold lacks causal powers, so too does the conjunction of $L$ and any core realizer.\(^\text{12}\)

All of this strikes me as correct, as far as it goes. I would add to McLaughlin’s line of thought that the causal status of total realizers was already very dubious, given CCP, even before the introduction of $L$, since total realizers must include (i) properties of the environment, (ii) historical properties, and (iii) versions of $T(x)$. None of (i-iii) seem to be causal properties; just as $L$ does not seem to be a causal property.

These considerations motivate a crucial denial: the denial that the total realizers upon which mental properties strongly supervene are causal properties. If we consider $P^*$ in Figure 1, it is very dubious that anything “causes” $P^*$. But if nothing causes $P^*$ then steps (3), (5), and (8) of the Supervenience Argument as presented in §3 are false. In particular, the “simple and natural” move made in step (3), in which we get $M$ to cause $M^*$ by having $M$ cause $P^*$, is shown

\(^\text{12}\) This is not to say that the core realizer itself lacks causal powers. A core realizer such as c-fiber firing may be brought into existence causally, via a causal process -- that is enough to assure that it is a causal property. But the conjunction of c-fiber firing and being such that the laws of nature hold cannot be brought into existence causally, via a causal process. This is why CCP has intuitive appeal.
to be misguided — misguided not because M’s causing of P* is preempted by P’s causing of P*, but rather because nothing causes P* at all.\textsuperscript{13} P* is a conjunctive property whose conjuncts are brought into existence in various ways. The only clear causal connection is between the core realizer conjunct of P and the core realizer conjunct of P* -- but M does not strongly supervene upon the core conjunct of P and M* does not strongly supervene upon the core conjunct of P*.

If we think about these matters in the more general terms of the exclusion problem and Bennett’s list, we see that we are led to deny Completeness. It is not the case that everything that happens in a physicalistic world has a physical cause. In particular, the instantiation of conjunctive properties such as P and P* do not have causes. That Completeness is false should not in itself be surprising.\textsuperscript{14} That the falsity of Completeness should so directly cause problems for the Supervenience Argument perhaps is surprising. We may of course make a simple change to Completeness to make it more plausible, altering it to, “Everything that happens that has a cause has a sufficient physical cause.” But this alteration highlights rather than obviates the problem for the Supervenience Argument: instantiations of total realizers are in the class of events in the physical world that happen but are not caused.

§5

We saw that in his 2005 presentation of the Supervenience Argument Kim wrote that the “simple and natural way” to ensure that M causes M* is to have M cause P* (Kim, 2005: 40).

\textsuperscript{13} For this reason, one cannot save the Supervenience Argument by moving to Kim’s “second completion” of the Supervenience Argument, the completion that does not rely upon P being a strong supervenience base for M (Kim, 2005: 44). The point is that even on the second completion, P* remains a strong supervenience base of M*, which is to say a total realizer of M*; that is all that is required in order to generate the difficulty. Throughout the current paper I have stuck with Kim’s first completion because it is more familiar, and because it allows me to raise issues concerning the standard causal-role playing explication of the so-called realization relation.

\textsuperscript{14} Kim himself authored a paper about uncaused events in the 1970s (Kim, 1993b).
We have ruled this out on the grounds that $P^*$ is too peculiar a property for that; it turns out that
*nothing* causes $P^*$. The next question we should ask -- the question that gets at the other of the
main points of this chapter -- is whether this observation does any good for the non-reductivist
position that the Supervenience Argument is meant to attack. In other words, we should now ask
whether all of this leaves open any way to save the idea that unreduced $M$s cause unreduced
$M^*$s. At first glance it seems that a way is available: we can save the causal efficacy of
unreduced mental properties by making the surprising claim that $M$ and $M^*$ are causal properties
even though their supervenience bases $P$ and $P^*$ are not. We can claim that $M$ causes $M^*$ even
though neither $P$ nor anything else causes $P^*$. If we can make this work then the non-reductive
physicalist might end up in good shape.

This line of thought does seem to have some initial plausibility. $M$ and $M^*$ supervene on
conjunctive properties but there is no reason to suppose that $M$ and $M^*$ are also conjunctive
properties. They, indeed, would seem *not* to be conjunctive properties. They are ordinary mental
properties such as pain and the belief that one is in pain. This suggest that transactions between
the $M$s are not just like transactions between the $P$s. The sort of non-causal dependence that $P^*$
has upon $P$ can be characterized via an appeal to the conjunctive nature of $P$ and $P^*$. But $M$ and
$M^*$ are not conjunctive, so the dependence that $M^*$ has upon $M$ cannot characterized in the same
way. So perhaps -- why not? -- the dependence of $M^*$ upon $M$ is causal even though the
dependence of $P^*$ upon $P$ is not.

Unfortunately, however, a big problem looms for the non-reductivist. It is very plausible
that *disjunctions* of non-causal properties are themselves non-causal properties. If total realizers
are non-causal properties, and if realized properties are disjunctions of total realizers, then it
seems that realized properties such as M and M* are non-causal properties. That realized properties are identical to disjunctions of their total realizers is the lesson of which Kim convinced many philosophers in his paper “Multiple Realization and the Metaphysics of Reduction.” There he wrote:

Pain is said to be a second-order property in that it is the property of having some property with a certain specification in terms of typical causes and effects and its relation to other mental properties; call this “specification H”. The point of MR [multiple realizability], on this view, is there is more than one property that meets specification H -- in fact, an open-ended set of such properties, it will be said. But pain itself, it is argued, is a more abstract but well-behaved property at a higher level, namely the property of having one of these properties meeting specification H. It should be clear why a position such as this is vulnerable to the questions that have been raised/ For the property of having P is exactly identical with P, and the property of having one of the properties P₁, P₂, . . . , Pₙ, is exactly identical with the disjunctive property P₁ v P₂ v . . . v Pₙ. On the assumption that Nh, Nr, and Nm are all the properties satisfying specification H, the property of having a property with H, namely pain, is none other than the disjunctive property, Nh v Nr v Nm! (Kim, 1993a: 323-24; italics original)

Many non-reductive physicalists found this line of argument convincing. Seeing no hope in distinguishing between a role property and the disjunction of its realizers, they tried to show that identifying role properties with disjunctions of total realizers wasn’t so bad, i.e. wasn’t so reductive.

Louise Antony explains the attractions of such a move in her 2003 paper “Who’s Afraid of Disjunctive Properties?”:

Suppose, for the moment, that we bite the bullet and simply identify presumably higher order mental properties with their lower-order disjunctive associates. If the latter count as “physical” properties, then we would be acceding to the weak reductionist
thesis that mental properties are physical properties. But so what? We can still maintain that mental properties are distinct from any of the disjunct properties—we can deny, for example, that mental properties are identical with neurological properties—and that’s the reductionist thesis that Kim is pushing. (Antony, 2003: 7; italics original)

The $N_h$, $N_r$, and $N_m$ in Kim’s Disjunction Argument have to be total realizers for the same reason that the $P$ and $P^*$ in his later Supervenience Argument have to be total realizers: $N_h$, $N_r$, and $N_m$ are stipulated to be unconditionally sufficient for the realized properties that they subvene. But given that, the problem should be obvious. If total realizers are not causal properties then neither are disjunctions of total realizers.

But what about the non-reductive functionalist who refuses to go along with Kim’s Disjunction Argument, who continues to distinguish between realized properties and the disjunctions of their total realizers? Jerry Fodor is such a functionalist, or anyway he has been willing to respond to Kim’s Disjunction Argument on behalf of such functionalist (Fodor, 1997). Unfortunately, in looking for a way to characterize realized properties that evaded Kim’s characterization of them as disjunctions of total realizers, Fodor characterized realized properties in another way that all too obviously also renders them as non-causal properties:

[I]t would be simply question begging of Kim to hold that being pain is the property of being one or other of pain's realizers. Functionalists claim that pains and the like are higher-order, relational properties that things have in virtue of the pattern of causal interactions that they (can or do) enter into. If so, then pains, though multiply based, are not disjunctive but MR. (Fodor, 1997: 153; italics in original)

Observe that Fodor calls realized properties, “properties that things have in virtue of the pattern of causal interactions that they (can or do) enter into.” But this renders them causally

\[^{15}\text{See footnote 6.}\]
irrelevant right off the bat. It cannot be the case that the property of being able to do something is the very property that enables the doing of it. To illustrate the point, the property of being able to run to second base very fast is not what propels a baserunner to second base -- rather, some complex physiological property of the baserunner does. The property of being able to remove lids from cans is not the property of a can opener that enables it to remove lids from cans -- this is done by the sharpness of the can opener’s blade, the force of its turning gears, etc.

The problem here seems to be quite general -- it is not as though Kim and Fodor are picking two ways of construing role properties out of some much larger variety of options. As long as we construe realization as causal role-playing it seems that we have to take realized properties to be causal role properties, and it further seems that causal role properties have to be taken in one of two ways: either as Kim or Fodor takes them. A causal role property is the property of having a property that plays a causal role, and that has to mean either (a) the property of having one of the total realizers, which if we believe Kim is the same thing as the disjunction of all of the total realizers, or (b) the property of being able to engage in the causal interactions that having one of the total realizers allows. Perhaps there is a third option, here, that of taking a role property to be the property of having one of the total realizers, without taking this to be the same thing as the disjunction of the total realizers. This third option would consist in taking option (a) and denying Kim’s addendum. But the burden is very much on the proponent of this
putative third option to show that the resulting property is a causal property even though the other two evidently are not.\textsuperscript{16}

The upshot of all of this is that we seem to be left, in Figure 1, with four non-causal properties. This is, to say the least, surprising. What I think it shows is that there is something wrong with this whole way of thinking about mental properties and their supervenience bases. What we need, perhaps, is a way of construing \( M \) and \( M^* \) such that they are causal properties even though the properties upon which they strongly supervene are not causal properties. That in turn suggests that we need a way of thinking about the realization relation, and thus the mind-body relation, that does not hinge on the notion of causal role-playing. This, it seems to me, is true, no matter what sort of physicalism we care to defend.

\section*{§6}

In this section I will more fully review three common sorts of response from non-reductive physicalists to Kim's Supervenience Argument, so that we may then see how my own response to Kim's argument compares. Responses of the first sort are called “compatibilist” and take issue with the principles of Exclusion and/or Nonoverdetermination upon which step (7) and (8) in Kim's argument rely. Responses of the second sort have it that the Supervenience Argument needs to assume a highly specific notion of causation in order to go through -- a

\textsuperscript{16} For example, one might try saying that the causal powers of a mental property are conferred by the core realizer while the remainder of the total realizer confers the enabling conditions that allow the powers of the core realizer to act. There is something right-sounding about this move, but as it stands it is unsatisfactory. What we need is a kind of causal property with which to identify mental properties. We cannot identify mental properties with core realizers because mental properties do not strongly supervene on core realizers, a necessary condition for identity. We cannot identify mental properties with total realizers because the latter are not causal properties (by CCP). Trying to combine the most helpful attributes of core and total realizers into a third property with which to identify mental properties might be the right idea, but the suggestion is the beginning of a research project, not the end of one. Note also that this would seem to be a non-reductive move, and so not the kind of move that might save the Supervenience Argument.
notion of causation that may be plausibly rejected. The third sort of response has it that the Supervenience Argument can’t be correct because if it is correct, it generalizes: its logic establishes that not just mental properties but all higher-level properties (biological, geological, chemical, etc.) are causally preempted by their physical subvenience bases. Let us go through these in turn.

**Compatibilism.** The compatibilist response to the exclusion problem dates back to Goldman’s (1969) response to Malcolm (1968). Compatibilism involves rejecting either Nonoverdetermination or Exclusion from the list of desiderata for non-reductive physicalism (to put the matter in terms of Bennett’s list). One can either reject Nonoverdetermination on the grounds that cases of mental causation exist and are genuine cases of overdetermination, or reject Exclusion (which, recall, says that there can be no double-causing except in cases of overdetermination) on the grounds that mental/physical double-causes are real but are not genuine cases of overdetermination because they are systematic. To put that differently, some compatibilists are happy to say that mental and physical causes genuinely overdetermine their common effects, while others would prefer to say that genuine overdetermination is objectionable, but that cases of mental and physical double causation are, because intimately linked, not cases of overdetermination as such. Versions of the compatibilist response to the exclusion problem have been put forward by many philosophers (c.f. Malcolm, 1968, Pereboom and Kornblith, 1991, Yablo, 1992, Yablo, 1997, Bennett, 2003).

When it comes to Kim’s Supervenience Argument, the compatibilist option translates to a rejection of (8) and possibly also (7), depending upon exactly what the compatibilist makes of the notions of Nonoverdetermination and Exclusion. Kim at first responded to this line of thought
by suggesting that M cannot be an additional sufficient cause for P* over and above the sufficient cause P because that would mean that there is a nearby possible world in which P does not occur but M does; in that world M causes P* in the absence of any physical cause for P*, thus violating Completeness (in other words violating the causal closure of the physical realm) (Kim, 1998: 44-45). Later, Kim changed his mind about this, observing that M could not occur without P in any case -- or at least, M could not occur without the occurrence of some physical supervenience base P’ such that P’ caused P* (Kim, 2005: 46-49). I add only that this response of Kim’s is further evidence that Kim takes M not merely to supervene upon P (or P’) but to be realized by P (or P’), where realization is understood as causal role-playing. It is only on this assumption that it is obvious that any supervenience base for M would cause a supervenience base for M*.

*Expanded sense of causation.* This is the least-frequently argued of the three sorts of response to Kim that I discuss, but it is the one to which Kim seems to regard his position as having the least satisfactory response. Sven Walter sums up this sort of objection nicely in a recent paper.

Kim’s real difficulty with overdetermination, the one which has apparently motivated his struggle against [non-reductive physicalism] throughout the past decades and which it has taken him years to state explicitly, has thus little to do with [the causal closure of the physical realm], the coincidental nature of typical cases of genuine overdetermination, or the apparent dispensability of overdetermining causes, and everything with his conviction that it is mandatory to think of mental-to-physical causation in terms of a productive/generative conception of causation. Here, however, a problem lurks. What legitimizes Kim’s appeal to his productive generative conception of causation? (Walter, 2008: 681)
According to this line of objection, Kim illicitly imports into his argument certain prejudices about how causation has to work -- prejudices that rule out mental causation as “parasitic” from the get-go, no matter what else we happen to think of firing-squad type cases of double causation. Those who accuse Kim of this prejudice point to passages such as the following, from *Mind in a Physical World*:

> If we understand the difference between genuine, productive and generative processes, on the one hand, and the noncausal regularities that are observed because they are parasitic on real causal processes, we are in a position to understand the picture recommended by [non-reductive physicalism]. In the case of supposed M-M* causation [see Figure 1], the situation is rather like a series of shadows cast by a moving car: there is no causal connection between the shadow of the car at one instant and its shadow an instant later, each being an effect of the moving car. The moving car represents a genuine causal process, but the series of shadows it casts, however regular and lawlike it may be, does not constitute a causal process. (Kim, 2000: 45)

Responding to a critique from Barry Loewer (Loewer, 2002) that this requirement flies in the face of accepted practice in physics, where causation is rarely spoken of and is not required to be the sort of generative process Kim describes, Kim replies with a striking lack of decisiveness:

> Loewer must be right -- or so I will assume -- when he says, as Russell did, that causality is not mentioned in “the fundamental laws and facts of physics.” He goes on to say:

> On the other hand, the productive conception of causation seems to involve facts that fail to supervene on the fundamental laws and facts. It seems to allow the existence of two universes that match perfectly in fundamental laws (of the sort posited by contemporary physics) and facts and yet differ in which events ‘produce’ which others.
I should admit at the outset that I am not in a good position to argue with Loewer, since his knowledge of physics vastly exceeds mine. But various thoughts occur to me when I reflect on his points.

My first thought is that if the situation is indeed as Loewer describes it and there is no productive causation anywhere, then there is no mental causation anywhere and the fear, common among philosophers and non-philosophers alike, that physicalism, or physics, takes away human agency, would be amply warranted . . .

Moreover, the fact that causality is not “mentioned” in the fundamental laws of physics, or that the word “cause (or any of its variants and synonyms) does not appear in the statements of these laws, should not be taken to show that the concept of cause is absent from physics. I think we should examine our understanding, or interpretation, of these laws, especially in explanatory contexts, and not become captives of the mathematical formalisms. (Kim, 2002: 675-76)

This seems unnecessarily conciliatory. Kim consistently compares supervenient causation to cases of illusory causation as with the shadow of the car, above, or with the images of apparent causal processes as reflected in a mirror (e.g. Kim, 2005: 37). All Kim needs is for causal processes to be more real than their reflections in order to rebut Loewer’s argument that physics makes no ruling about which causal processes are legitimate. That is to say, surely Loewer doesn’t want to claim that physics makes no distinction between causal processes and their mirrored reflections. Kim claims that the distinction between subvenient and supervenient causation is like that. Perhaps Kim is wrong to say that the distinction between subvenient and supervenient causation is like that, but Loewer does not deny that it is.

What is more interesting is that all of this shows that Kim is not only relying upon the Supervenience/Exclusion Argument to create problems for mental causation, even when it superficially appears that he is. The exclusion problem, after all, takes it for granted that mental
causation is or can be *genuine* causation -- *i.e. not* like a reflection or shadow of actual causation. As Bennett points out, the Exclusion Argument’s refusal to stack the deck against mental causation by dismissing the genuineness of mental causation out of hand is what makes the Exclusion Argument so intriguing (Bennett, 2003: 471). Therefore Kim’s occasional lapses away from the Exclusion Argument into assertions that mental causation is simply not genuine causation anyway, are not to the same point. For that reason, responses to Kim’s claims that mental causation is not genuine causation are not responses to Kim’s Supervenience Argument (even if everyone involved takes them to be).

*Generalization.* The objection that the Supervenience Argument cannot be correct because if it is correct it generalizes has been presented by several philosophers, among them Noordof (1999), Gillett and Rives (2001), and Bontly (2002). The idea is that the same logic that Kim applies to mental properties to show that they are causally superfluous if unreduced can equally well be applied to all higher-level properties, for example chemical, geological, economic, and meteorological properties. If we consider Figure 1, this line of thought has it that mental properties that makes them uniquely suited to be the Ms; any other kind of special- or folk-scientific property could just as well be slotted in for the Ms. But that means that causation between any special or folk-scientific properties is just as illusory as the casual relation between M and M* taken as mental properties. But, friends of the generalization argument tell us, this is absurd. It cannot be the case that all special- and folk-scientific properties are causally superfluous.

---

17 This makes problematic Kim’s claim that the Exclusion Problem is “motivated” (Kim, 2005: 32) by cases such as the mirrored causal processes of Edward’s Dictum. (Kim, 2005: 32-38) Perhaps for Kim such cases are what motivates the Exclusion Problem, but they should not so motivate anyone else, since they and the exclusion problem have nothing to do with each other, and indeed are mutually exclusive.
Note that this line of response to the Supervenience Argument does not purport to tell us what is wrong with the Supervenience Argument, only that something must be wrong with it.

Bontly sums-up the generalization argument in his 2002 paper:

Since the premises [of the Supervenience Argument] look to be perfectly general theses about supervenient events or properties, the argument should generalize to threaten biological causation, chemical causation -- i.e., all causation at the macro-level. Widespread ‘macro-epiphenomenalism’ would seem to ensue.

On its face, however, that result appears to be patently absurd. First of all, it seems to mean the complete rejection of causal explanations in the special sciences: not only do thoughts not cause behavior, but likewise it appears to follow that smoking is not a cause of cancer, and that genes do not causally contribute to phenotype. . . Any philosophical thesis that conflicts with so much successful science must surely be viewed with suspicion, to say the very least, so the friend of the supervenience argument needs some way to block the the threatened reductio. (Bontly, 2002: 78-79)

Kim seems to have offered two sorts of response to the generalization argument. The first sort of response attempts to block the generalization and thereby the reductio (Kim, 2000: 80-87). To put this in terms of Figure 1 again, Kim tries to show that mental states really are specially suited to be the Ms; suited in a way that special science properties are not. To show this, Kim makes use of a distinction between higher-order and higher-level properties. To see how this is supposed to work, think of physical properties as first-order properties of objects at the lowest of mereological levels; in other words, think of physical properties as first-order properties of objects that are not composed. The properties studied by the other sciences will then be “non-physical” properties in the sense that they will either be not first-order, or not had by objects that are not composed, or both.
Kim insists that while higher-order properties can have only causal powers that are already possessed by those properties’ lower-order realizers, higher-level properties may have causal powers that are genuinely novel -- they may have powers that are not had by properties of objects at lower mereological levels. To put that last point another way, composed objects may have properties that bestow causal powers that are not had by any of the properties of their proper parts:

These considerations suffice to show that the generalization argument does not have the full generality its supporters attribute to it. In particular, the exclusion-based worries about mental causation do not generalize across micro-macro levels. Consider Block’s worries about second-order properties[18]: here the main worry is that the causal role of a second-order property is threatened with preemption by its first-order realizers. As we noted, second-order properties and their realizers are all at the same level in the micro-macro hierarchy, and the causal worries about second-order properties are _intralevel_ worries and do not cut across micro-macro boundaries. The exclusion-based arguments . . . therefore do not generalize across micro-macro levels . . .

_Macroproperties can, and in general do, have their own causal powers, powers that go beyond the causal powers of their micro-constituents._ (Kim, 1998: 84-85; italics original)

Since I intend to show that the Supervenience Argument does not work, it should be predictable that I do not think it generalizes. However, even on the assumption that the Supervenience Argument does work, this response from Kim seems to be entirely beside the point. The Supervenience Argument, after all, does not hang from M (see Figure 1) having the same or different causal powers from P. What matters is that M’s causal powers (whatever they

---

18 Kim is referring to Block’s paper “Can the Mind Change the World?” (Block, 1990) in which Block expresses worries about the causal efficacy of second-order properties.
are preempted by the powers of P. According to the Supervenience Argument M can cause M* only by causing P*, but M can’t cause P* because P causes P* and P* can’t be doubly caused. Nothing in that line of reasoning is directly thwarted by saying M has causal powers “that go beyond the causal powers” of P. The reason M can’t cause M* directly is not that it lacks the power to do so; presumably it has the power to do so (if it didn’t there would be no point to the Supervenience Argument). The reason M can’t cause M* directly is simply that M*’s existence is already assured by the presence of its own supervenience base, P*. Therefore, the business about the difference between the causal powers of higher-level and higher-order properties is of no consequence.

Kim should probably just bite the bullet and accept the consequences of the generalization argument -- should probably just accede that all of the genuine causal goings-on in the world are goings-on at the physical level, because the causal transactions at that level (and order) pre-empt all the rest. And that is just what Kim does say in his second sort of response to the generalization argument, presented in his 2005 book *Physicalism or Something Near Enough*.

Indeed, the supervenience argument may be generalizable, but all that would show is that if there is biological causation, biological properties are, or are reducible to, physical or physico-chemical properties; it does not show that biological causation does not exist. The epiphenomenalism brunt of the [supervenience] argument is avoided if one is prepared, and able, to choose the reductionist branch of the dilemma. (Kim, 2005: 55)

This response to the generalization argument is superior to Kim’s first attempt because it presses the point that Kim was never out to show that mental properties don’t cause anything. As he sometimes tries to remind his critics, he is not an epiphenomenalist. (Kim, 2005: 54) He is out to save the causal efficacy of the mental in a physical world, just as are his interlocutors. Kim
simply has a different and more drastic idea about how far one must be prepared to go in order to
do that. The “bite the bullet” response to the generalization argument does a better job of
keeping to that spirit.
Chapter Four

Two Kinds of Role Property

§1

According to a popular view in the philosophy of mind, mental properties are realized properties, and realized properties are causal role properties. According to this view, to say that neurological property N of the brain realizes some mental property M of the brain is to say, inter alia, both that N stands in the causal role-playing relation to M and that M is a causal role property. We may call this view of the nature of mental properties ‘causal role functionalism’. According to causal role functionalism, then, the mind-body relation is the realization relation, and the realization relation is the causal role-playing relation.

In this chapter I discuss a problem with the characteristic move of causal role functionalism: taking realization to be causal role-playing. I argue that while realization is intended to be a one-one relation, such that each instantiation of realization relates a single realizer property to a single realized property, the causal role-playing relation is evidently a one-many relation, such that each of its instantiation relates a single role-playing property to at least two distinct role properties. If this is correct, then the causal role-playing relation is not a good candidate for being the realization relation, and so by extension causal role functionalism is not a good candidate for being the right way to think about functionalism.

What is obviously crucial to my argument is the claim that the causal role-playing relation is a relation between a role-playing property and at least two role properties. To make

1 A paper closely based on this chapter is forthcoming in Philosophia.
good on this claim, I need to show that there is more than one second-order property that an
object necessarily has just in case the object has a first-order property that plays a given causal
role. If this can shown, then causal role functionalists will have the initial problem (though this
will not be the only problem) of determining to which sort of property a philosopher is referring
when the philosopher identifies mental properties with role properties.

I will proceed as follows. In §2 I will present initial reasons to think that there is more
than one sort of second-order “role property” that an object necessarily has just in case it has a
first-order property that plays a causal role. In sections §3-5 I will examine a specific recent
debate in the philosophy of mind -- the debate over Jaegwon Kim’s Disjunction Argument -- in
order to show how this plurality of role properties works to create confusions. In §6 I will take a
step back from the specific example of the Disjunction Argument to consider what all of his
means for causal role functionalism generally.

§2

I begin with what I take to be an unexceptionable formulation of the standard definition
of causal role properties, followed by a brief collection of examples taken almost at random from
the literature. First, my formulation; I will call it the Generic Role Property Definition:

\[(\text{GRP}) \quad \text{Role property } Z =_{\text{def}} \text{the second-order property of having some}
\quad \text{first-order property } T \text{ that plays a certain causal role } C.\]

Now some typical examples from the literature:

A second-order property is the property of having some first-order
property that satisfies a certain condition. And the condition, at
least in the case of first-order properties that realize functional
properties, is said to be the having of a certain causal role.
(Shoemaker, 2007, p. 11)
F is a second-order property over set B of base (or first-order) properties iff F is the property of having some property P in B such that D(P), where D specifies a condition on members of B . . . We may now explain functional properties over B as those second-order properties over B whose specification D involves the causal/nomic relation. (Kim, 1998, p. 20)

Fness = the property of having an F-ish property. Because [this] represents Fness as the property of having a property that plays a causal role, let's call it the role theory of Fness. (Yablo, 1995, p. 481)

The property M = the property of having some (first-order) property φ, such that being in I[put] causes one to have φ and having φ causes one to go into O[utput]. (David, 1997, p. 134)

Such examples could be multiplied, but these four are enough to convey the commonalities that GRP is intended to capture: the way in which the base property, the role, and the role property are typically described in such definitions. The base property is typically described as ‘some property’ or ‘a property’ that does something describable as playing or having a causal role. The role itself, as we can see from Kim’s and David’s formulations, need not be called a ‘role’ or even explicitly mentioned -- what matters is that each of the above definitions includes a reference to a specification of the causal interactions into which the base property enters. The role property is in each case defined as the property of having a base property (as just described) play a causal role (as just described). In sum, standard definitions of causal role properties -- i.e. versions of the generic definition -- define role properties as the kind of second-

---

2 The base property is also typically said to be a ‘first-order property’. This is stated in my version of the generic definition and will be assumed in what follows, but the generic definition may be stated without that requirement. Some philosophers think that second-order properties can play the causal roles of third-order properties, and so on up the ladder of orders, in the just the same way that first-order properties play the role of second-order properties. Whether one finds this plausible will depend on what one thinks of the causal efficacy of second-and-higher order properties. In any case, what I say in this chapter applies to first-order base properties and second-order role properties; but I take it that what I say is mutatis mutandis upwardly applicable.
order property that is had by objects when and only when those objects have first-order
properties that enter into or anyway are able to enter into a specifiable list of causal interactions.
That is the common definitional coin that my GRP is meant to capture.

The problem is that GRP is ambiguous. There are at least two kinds of second-order
property that satisfy the definition. GRP could be taken to define the second-order property of
having the disjunction of all of the first-order total realizers that satisfy a given role; or it could
be taken to define the second-order property of being able to enter into the pattern of causal
interactions that having one of the total realizers allows.\(^3\) To put that a bit less technically, GRP
could be taken to define either the property of having a role-player or the property of having the
role played. Both readings of GRP are present in the literature and they lead to two substantially
different takes on the nature of what is supposed to be a single kind of role property. Once we
disambiguate GRP we get two definitions that I will call ‘(D-A)’ and ‘(D-B)’. They define two
kinds of role property that I will call ‘A-type role properties’ and ‘B-type role properties,’
examples of which I will call ‘A’ and ‘B’ respectively.

\[
\text{(D-A) Role property } A =_{df} \text{ the second-order property of having the}
\text{disjunction of first-order properties } T_1, T_2, \ldots, T_n \text{ where the}
\text{set } \{T_1, T_2, \ldots, T_n\} \text{ is the set of possible total realizers that}
\text{play causal role } C.
\]

\(^3\) I say that GRP is ambiguously satisfied by “at least” two different kinds of role or functional property; but it is
obvious enough that there are more. Amongst functionalists who take role properties to be disjunctions of possible
role players there are those who take role properties to be identical to the disjunction of all of the metaphysically
possible realizers (Melnyk, 2003; Shoemaker, 1981), and those who take them to be identical to the disjunction of
just the physically possible realizers (Antony, 2003; Bennett, 2008). I take it though that both disjunctive properties
are had by any object in the actual world (or any physically possible world) that has either one of them. Given this,
there are at least three kinds of role property: the type that satisfies D-B and two types that satisfy D-A.
Role property \( B =_\alpha \) the second-order property of having \( C \) played (by any first-order property).\(^4\)

Note the occurrence of the phrase ‘total realizer’ in D-A. I adopt Sydney Shoemaker’s distinction between total and core realizer properties (Shoemaker, 1981, 2007) because the distinction is necessary to state D-A correctly. A core realizer is the base property associated with the functional property it realizes, as with \( c\text{-fiber-firing} \) and \( pain \). The total realizer is the core realizer conjoined with what Shoemaker calls ‘the surround’ (2007, p. 21). The surround is the stable structural property of an entity that allows the core realizer to have the kinds of causes and effects that go into playing the role. In order to instantiate a role property such as \( pain \) it is not enough to instantiate the core realizer; the total realizer must be instantiated. This much should be familiar: putting a firing c-fiber into a Martian or other inappropriately structured organism will not result in pain on the functionalist conception of pain, because the firing c-fiber won’t be hooked up in the right ways. That is to say: the Martian lacks a core-appropriate surround. Therefore, A-type role properties are properties of having the disjunction of the possible total realizers, not (merely) the possible core realizers. Having said all of that, I should stress that I do \textit{not} take the distinction between (D-A) and (D-B) to hang from a distinction between total and core realizers.

\(^4\) Note the occurrence of the phrase ‘total realizer’ in D-A. I adopt Sydney Shoemaker’s distinction between total and core realizer properties (Shoemaker, 1981, 2007) because the distinction is necessary to state D-A correctly. A core realizer is the base property associated with the functional property it realizes, as with \( c\text{-fiber-firing} \) and \( pain \). The total realizer is the core realizer conjoined with what Shoemaker calls ‘the surround’ (2007, p. 21). The surround is the stable structural property of an entity that allows the core realizer to have the kinds of causes and effects that go into playing the role. In order to instantiate a role property such as \( pain \) it is not enough to instantiate the core realizer; the total realizer must be instantiated. This much should be familiar: putting a firing c-fiber into a Martian or other inappropriately structured organism will not result in pain on the functionalist conception of pain, because the firing c-fiber won’t be hooked up in the right ways. That is to say: the Martian lacks a core-appropriate surround. Therefore, A-type role properties are properties of having the disjunction of the possible total realizers, not (merely) the possible core realizers. Having said all of that, I should stress that I do \textit{not} take the distinction between (D-A) and (D-B) to hang from a distinction between total and core realizers.
The distinction between A- and B-type role properties will be explored over the course of this chapter; but as a first pass it is similar to the distinction between, on the one hand, having an employee who fills a job position, and on the other hand, having the job position filled. Another illustration: it is similar to the distinction between, on the one hand, a fast car’s property of being able to go from 0 to 100 kilometers-per-hour in five seconds, and on the other hand, that car’s property of being one of the types of car that, as it happens, can accelerate that quickly: being a Bugatti Veyron or a Ferrari Testerossa or a Porsche 911 or a Chevrolet Corvette or . . . (here we make a disjunctive list of all of the actual and all of the possible but non-actual car designs that can go from 0 to 100 kph in 5 seconds). These properties are co-extensive, but they are not, it seems obvious, the same.\footnote{They have different truth-makers, for example. What makes it true that a car is a Bugatti Veyron or a Ferrari Testerossa or . . . is not the same thing as what makes it true that a car can go from 0 to 100 kph in 5 seconds.}

Analogies with employees and automobiles aside, the difference between A- and B-type role properties can be seen in the different philosophical uses to which they are put, and the confusions which result from not taking note of the difference between them.

I will proceed in this chapter as follows. As an example of the confusions that can be caused by a failure to note the distinction between A- and B-type role properties, in the next three sections I examine the debate over Jaegwon Kim’s Disjunction Argument as presented in his agenda-setting paper ‘Multiple Realization and the Metaphysics of Reduction’ (Kim, 1993). In §3 I will review Kim’s argument. In §4 I will examine Jerry Fodor’s (Fodor, 1997) response. In §5 I will examine responses by Ned Block (Block, 1997), Lenny Clapp (Clapp, 2001) and others. What we will see is that both Kim and Block take GRP to be read as D-A while Fodor takes GRP to be read as D-B. Seeing this is crucial to understanding why Fodor and Block each reply to
Kim as they do, and it is also crucial for understanding Block’s (mis)reading of Fodor’s response to Kim.

Although I take improved understanding of the debate over the Disjunction Argument to be a laudable goal in and of itself, such is not my primary aim. The relevance of the distinction between A- and B-type role properties to functionalist metaphysics is quite general. Wherever causal roles are employed to explicate the realization relation, the duality of role properties will be there to cause mischief. Thus, I intend for §3-5 to serve dual purposes. §3 presents an illustrative example of how the conflation of A-type role properties with B-type role properties happens. §4 explains B-type role properties: what they are, how they work, and to what philosophical work they may be put. §5 is intended similarly to explain A-type role properties.

Finally, in §6 I will more explicitly consider the problem presented by the ambiguity of GRP to functionalist metaphysics generally. In a nutshell, the problem is that it is not as though one of D-A or D-B correctly defines the one real sort of role property while the other fails to define anything. If A-type role properties exist then so do B-type role properties, and vice versa. The issue is whether this duality of role properties results in a sort of metaphysics with which a functionalist about mental states ought to be comfortable.6

6 Since the discussion to follow crucially involves the views of Jaegwon Kim as presented in ‘Multiple Realization and the Metaphysics of Reduction’, I should note that Kim wrote this chapter before he began insisting upon a hard distinction between higher-level and higher-order properties. (He made the distinction most forcefully in the later Mind in a Physical World (Kim, 1998, pp. 80-87).) The responses to Kim’s 1992 paper from other philosophers that I consider also do not take this hard distinction into account. If we make the distinction between higher-order and higher-level properties then the discussion to follow applies to higher-order role properties, not higher-level role properties. (I am not even sure what higher-level role properties would be, when the distinction between orders and levels is made the way the later Kim makes it.) This is fine, since by common agreement among functionalists psychological properties are what Kim calls higher-order properties, and it is psychological properties with which I will be most directly concerned.
In this section we will look at Jaegwon Kim’s Disjunction Argument in order to set the stage for our discussion of responses to it by Fodor, Block and others. We will also see in this section an example of the misuse by Kim of a version of the Generic Role Property Definition. Kim did not, in his construction of the Disjunction Argument, appreciate that the definition is ambiguous; this will create trouble down the road.

Let us begin by reminding ourselves of the view that the Disjunction Argument is intended to undermine: the view, staked-out by Jerry Fodor in his classic 1974 paper ‘Special Sciences’, that higher-level special science properties are nomically autonomous with respect to properties studied by lower level sciences. Fodor (1974) argued that a special science property, say, the psychological property pain, is autonomous if and only if both that special science property is nomically homogenous and the set of lower-level base kinds that realize it is nomically heterogeneous. Nomically homogenous scientific properties obey a tidy set of natural laws. The fact that instances of a causally homogeneous property have behaved in a characteristic way in the past is evidence that the next instance will behave in that characteristic way, too. On the other hand, nomically heterogeneous properties are, Fodor said, ‘wildly disjunctive’ (Fodor, 1974, p. 124); the causal behavior of past instances of a wildly disjunctive property is no evidence that the next instance of that property will behave in a similar way.

Fodor claimed, uncontroversially at the time, that causally homogenous higher-level properties such as pain supervene on collections of wildly disjunctive lower-level properties. Different instances of pain may have, as realizers, instances of lower level properties that have nothing in common from the point of view of their proprietary special science. Organisms that
from a biological point of view are nothing alike may nevertheless have the same psychological
properties.\footnote{The empirical version of this claim, that psychological properties either could easily be or actually have been
shown to have multiple neurobiological realizations in different species or in the same species at different times, was
later challenged, notably by Bechtel and Mundale (1999). More philosophical, as opposed to empirical, reasons for
skepticism about the prevalence of multiple realizations were presented by, for example, Shapiro (2000) and Polger
(2004).} Science can’t study the natural laws that govern all these biological realizers of pain
because there are no such natural laws. Though one could force hideously complex disjunctive
laws obeyed by all realizers of pain into one’s lower-level theory, such laws would be arbitrary,
unreliable and, in a deep sense, fake. Therefore, Fodor concluded, the special sciences need not
fear reduction to lower-level sciences, ultimately to physics.

Kim’s Disjunction Argument turns Fodor’s argument on its head. Kim argued that the
causal powers of each instance of a given higher-level property come from nowhere other than
instances of subvening lower-level properties. So if the disjunction of the lower-level properties
that variously realize a given higher-level property is nomically wild -- that is, if the various
lower-level properties have no causal powers in common -- then surely the instances of the
allegedly tidy higher-level property are just as causally disparate as are instances of its various
realizers. On the other hand, if the lower-level realizing properties are not causally disparate, if
they do have causal powers in common, then Fodor’s argument provides no reason to believe
that the higher-level property is distinct from, or autonomous with respect to, the disjunction of
these lower-level realizers. So special science properties either are not natural kinds or else are
reducible. As Block writes, ‘The power of Kim’s argument is that reductionism beats the Many
Levels view and multiple realizability either way’ (Block, 1997: 114).
Now let us have a closer look at how Kim’s argument is supposed to work. I contend that it is crucial to the Disjunction Argument that it states a version of the generic role property definition and then disambiguates it as D-A. I quote at the crucial passage at some length.

Pain is said to be a second-order property in that it is the property of having some property with a certain specification in terms of its typical causes and effects and its relation to other mental properties; call this “specification \( H \)”. The point of MR [multiple realizability], on this view, is that there is more than one property that meets specification \( H \) – in fact, an open-ended set of such properties, it will be said. But pain itself, it is argued, is a more abstract but well-behaved property at a higher level, namely the property of having one of these properties meeting specification \( H \). It should be clear why a position like this is vulnerable to the questions that have been raised. For the property of having \( P \) is exactly identical with \( P \), and the property of having one of the properties, \( P_1, P_2, \ldots, P_n \), is exactly identical to with the disjunctive property, \( P_1 \lor P_2 \lor \ldots \lor P_n \). On the assumption that \( N_h, N_r, \) and \( N_m \) are all the properties satisfying specification \( H \), the property of having a property with \( H \), namely pain, is none other than the property of having either \( N_h \) or \( N_r \) or \( N_m \) – namely the disjunctive property \( N_h \lor N_r \lor N_m \)!

We cannot hide the disjunctive character of pain behind the second-order expression, “the property of having a property with specification \( H \)”. Thus, on the construal of mental properties as second-order properties, mental properties will in general turn out to be disjunctions of their physical realization bases. It is difficult to see how one could have it both ways – that is, to castigate \( N_h \lor N_r \lor N_m \) as unacceptably disjunctive while insisting on the integrity of pain as a scientific kind. (Kim, 1993, pp. 323-324; second italics added)

In this passage Kim first describes his opponents’ view, then writes a transition sentence (‘It should be clear why a position like this is vulnerable . . .’), and then argues that his opponents’ view doesn’t work. Prior to the transition sentence, Kim twice describes role-properties by using a version of GRP. First, ‘Pain is said to be a second-order property in that it is the property of having some property with a certain specification in terms of its typical causes
and effects and its relation to other mental properties’. Second, ‘pain itself, it is argued, is a
more abstract but well-behaved property at a higher level, namely the property of having one of
these properties meeting specification $H$’.

After the transition sentence, Kim twice describes role-properties with a version of D-A. First, he describes a role property as ‘the property of having one of the properties, $P_1$, $P_2$, . . . ,
$P_n$’. Second he writes, ‘On the assumption that $N_h$, $N_r$, and $N_m$ are all the properties satisfying
specification $H$, the property of having a property with $H$, namely pain, is none other than the
property of having either $N_h$ or $N_r$ or $N_m$’. Thus Kim more-or-less explicitly states that every
property of having a property with $H$ is identical to the property of having one of $P_1$, $P_2$, . . . , $P_n$.

That premise is critical to the argument. Let us set it out clearly and give it a designation.

(U1) Every property of having a property that meets a certain
specification $H$ is identical to the property of having one of
$P_1$, $P_2$, . . . , $P_n$, where $P_1$, $P_2$, . . . , $P_n$ is an exhaustive list
of properties that meet $H$.

From there, Kim claims that the property he has just described (via U1) is identical to a
disjunctive property. That is, he claims that the property of having $N_h$, $N_r$, or $N_m$ is identical to
the disjunctive property $N_h \lor N_r \lor N_m$.

(D1) Every property of having a member of a list of properties is
identical to the disjunctive property that is formed by
disjoining all of the members of the list.

What I want to stress is that there two identity claims at work here. Both are needed in
order to establish that all role properties are disjunctive properties -- that all properties that
satisfy GRP are properties that satisfy D-A. A philosopher who wishes to identify mental
properties with role properties but wishes to deny that mental properties are disjunctive
properties can try denying either or both of U1 and D1.
I believe that what Kim’s argument does most effectively is defend D1. Kim’s argument does not seem to me to do anything to defend U1. As it happens, after Kim presented his argument most non-reductive functionalists accepted his overall conclusion that role properties are identical to disjunctive properties. I believe that that is a good sign that prior to Kim’s argument most non-reductive functionalists were implicitly assuming that the denial of D1 does all of the work in separating role properties from disjunctive properties. That is D1 did all the work in separating what I am calling B-type role properties from A-type role properties. After Kim refuted D1 (to the satisfaction of most non-reductive functionalists), non-reductive functionalists looked for ways to live with the result that mental properties were disjunctive properties. They looked for ways to accept Kim’s overall conclusion while denying that it has as much reductive import as Kim thought. We will see that kind of strategy in §5, employed by a group I refer to as “the non-reductive disjunctivists of the 90s,” though they continue into the 2000s.

First though, we should note that it is still open to the non-reductivist to resist Kim’s overall conclusion by denying U1. This option did not seem to occur to many non-reductivists in the 1990s or since; but it did, I think, occur to Fodor. I believe that he put this option to use in his 1997 response to Kim’s argument, the response to which we now turn. As we’ll see, in constructing this response to Kim, Fodor abandoned central aspects of his earlier, 1974 view.
In his 1997 response to Kim, Fodor wrote of the dispute that I am here characterizing as a
dispute between those who identify mental properties with A-type role properties and those who
identify them with B-type role properties that it constitutes ‘a sort of polemical standoff’:

This is, as I say, a sort of polemical standoff. The functionalist [i.e. the advocate of B-type role properties] assumes that there are laws about pains ‘as such’; so he infers that, though pain is multiply based, it is not (merely) disjunctive. So he infers that pain is unlike jade in the respects that are relevant to the question of projectibility. Kim, going the other way around, assumes that pain is (merely) disjunctive, hence that it is relevantly similar to jade, and hence that there aren’t any laws about pain. The real issue – the one that Kim’s appeal to the jade example begs – is whether there is a difference between a multiply based property’s being disjunctive and its being MR; and, if so, what the difference is and whether it matters to projectibility. (Fodor, 1997, pp. 153-154; italics original)

According to Fodor, non-reductive functionalists take realized properties to be nomically homogenous, and from there argue as follows. Since realized properties are nomically homogenous, it follows they cannot be identical to properties that are nomically heterogenous. But disjunctive properties are nomically heterogenous. Therefore, realized properties cannot be identical to disjunctive properties. Kim, on the other hand, begins with the observation that disjunctive properties are in danger of being nomically heterogenous, makes the assumption that realized properties are identical disjunctive properties, and concludes that realized properties are nomically heterogenous if disjunctive properties are.

Fodor proposes in his 1997 paper to get around the polemical standoff by pointedly not caring whether the disjunction of a realized property’s realizers forms a causally heterogeneous or a causally homogenous disjunctive property. He intends to show that a realized property is simply not the same thing as its corresponding disjunctive property, and that the nomic status of
the former does not stand or fall with the nomic status of the latter. While Fodor’s 1997 paper is not his clearest, I think the best way to read it is as arguing for this tactic by denying U1 and not caring about D1.

Fodor does not dispute that (merely) disjunctive properties are, or are in danger of being, nomically heterogenous. He also does not deny that any object that has a realized property also has the merely disjunctive property that is the disjunction of the realized property’s possible realizers. What Fodor denies is that the realized property is identical to the disjunctive property. From there, he argues that the nomic heterogeneity of the disjunctive property, should it be such, does not establish the nomic heterogeneity of the realized property.

This was an unusual line of argument. As I noted earlier, at the time that Fodor mounted this defense of non-reductive functionalism, most non-reductive functionalists were conceding to Kim the point that realized properties were identical to the disjunctions of their realizers. Fodor was nearly alone in resisting the identification. I think the right way to read this history is that most non-reductive functionalists accepted D1 and thought that was sufficient to force the disjunctive move. Fodor, on the other hand, denied a prior move that few other people noticed or took to be in play: he denied U1.

But if one denies U1, what kind of property is one taking realized properties to be? Fodor writes:

[I]t would be simply question begging of Kim to hold that pain is the property of being one or another of pain’s realizers. Functionalists claim that pains and the like are higher-order, relational properties that things have in virtue of the pattern of causal interactions that they can (or do) enter into. If so, then

---

8 I think that one can read parts of Antony and Levine’s 1997 paper as denying the identity of realized properties with the disjunctions of their realizers (Antony & Levine, 1997). But the denial is not as emphatic as Fodor’s and, in any event, by 2003 Antony, at least, was conceding the identity (Antony, 2003).
Fodor takes realized properties to be “properties that things have in virtue of the pattern of causal interactions that they can (or do) enter into.” Thus, a realized property is the property of being able causally to do something. Or, to put that differently, the property of having a causal role satisfied or played. This kind of property is what I am calling a B-type role property.

Fodor’s response to Kim’s Disjunction Argument is essentially an admonition to attend to the difference between two sorts of property. Fodor does not, of course, use my terms ‘A-type’ and ‘B-type’ for the properties he distinguishes, and in fact the distinction that he draws is not strictly parallel to my distinction between A- and B-type role properties. Before we get to that, though, it’s important to see how Fodor changes his 1974 view in order to combat Kim’s argument.

Recall that according to Fodor’s 1974 view, a realized property $G$ is autonomous with respect to its lower level realizer properties $P_1, P_2, \ldots, P_n$ if and only if both $G$ is a natural kind and the disjunction of $G$’s lower-level realizer properties $P_1, P_2, \ldots, P_n$ is not. The 1974 view thus upheld the following two criteria for MR (what follows is my wording):

\begin{equation}
\text{(MRP)} \quad \text{A higher-level property } G \text{ is MR with respect to its lower-level base properties } P_1, P_2, \ldots, P_n \text{ iff (i) } G \text{ is a natural kind, which is to say projectible, which is to say nomically homogeneous, and (ii) the disjunctive property } P_1 \lor P_2 \lor \ldots \lor P_n \text{ is not a natural kind, which is to say not projectible, which is to say not nomically homogeneous.}
\end{equation}

With that in mind, consider the following passage from Fodor’s 1997 paper:

We are told, Kim tells us, that jade “is not a mineral kind, contrary to what was once believed; rather, jade is composed of two distinct minerals with similar molecular structure, jadeite and nephrite.” [(Kim, 1993, p. 319)] Kim thinks that, because of these
facts about jadeite and nephrite, jade is paradigmatically MR. Kim thinks that since “jade” is paradigmatically MR, it is ipso facto unprojectible. And, a fortiori, (3) isn’t a law.

3. Jade is green.

*I don’t actually care much whether (3) is a law or even whether “jade” is projectible. But I am going to deny that jade is paradigmatically MR . . .* (Fodor, 1997: 151; final italics added)

Fodor is here stating that jade is not a multiply realized kind whether or not the property *being jade* meets criterion (i) of MRP. Let us play out what this means. Say for the sake of argument that *being jade* meets criterion (i) of MRP. Then in order for jade to count as a multiply realized kind all we need is for the disjunctive property *being jadeite or nephrite* to meet (ii). But elsewhere in the 1997 paper Fodor grants (what is obviously true anyway) that the disjunction of *being jadeite* and *being nephrite* meets condition (ii). So Fodor is adopting a view according to which a kind or property could meet *both* of the criteria stated in MRP and still not be MR. What is the new view?

Well, to be repetitive about it, the new view constitutively involves the denial of U1. Fodor makes a distinction between the property of having one of a multiply realized property’s realizers, and the realized property itself. The former is (by D1, which Fodor does not deny) an A-type role property. The latter is a B-type role property -- the property of have a causal role played. In effect, Fodor is claiming that *being jade* is what I am calling an A-type role property while *pain* is what I am calling a B-type role property, and that it would be ‘simply question begging’ of Kim to insist that *pain* is an A-type role property.

It is crucial to note that Fodor does not deny the existence of the A-type role property that corresponds to the B-type role-property that is pain. Pain, of course, has a set of base realizer
properties, and there is such a thing as the disjunctive property that is the disjunction of those base realizer properties. But that property, the A-type role property that anything has just in case it has one of pain’s realizers, is not pain. Fodor explicitly distinguishes between them here:

> Functionalists think that there are laws about pain, but they don’t think that there are laws about jade, and they also don’t think that there are laws about the metaphysically open disjunctive realizer of pain. (Fodor, 1997, p. 156; italics original)

The ‘metaphysically open disjunctive realizer of pain’ is the property one gets if one disjoins the members of the set of pain’s realizers. In saying that there are laws about pain whether or not there are laws about the disjunction of pain’s realizers, Fodor is claiming that they are not the same thing.

Next we turn to Ned Block’s response to Kim. Block adopts a strategy for responding to Kim’s Disjunction Argument that has become common among functionalist-minded philosophers who accept Kim’s disambiguation of GRP as D-A. Block argues that although mental properties are identical to disjunctive properties, just as Kim says they are, still the disjunctive properties in questions aren’t so wildly disjunctive after all.

§5

At the start of §4 I quoted a passage from (Fodor, 1997) in which Fodor implied that most functionalists would not accept Kim’s claim that, according to functionalism, mental properties are to be identified with what I am calling A-type role properties. It turns out that Fodor underestimated the willingness of many functionalists to go along with Kim’s assumptions. After Kim presented his disjunction argument a number of functionalist philosophers who wished to defend (at least some sort of) non-reductivism about mental properties agreed with
Kim that GRP was to be interpreted as D-A. They were persuaded by Kim’s claim D1 that the property of having a member of a list of properties was identical to the disjunction of the members of the list. They went to work on the nature of disjunctive properties in the hopes of finding a way to identify mental properties with disjunctive properties that was not odiously reductive. If successful, they thought that this would suffice to save psychological and other special-science properties from Kim’s reductivist attack.

Antony provided a clear statement of the attractions of this approach:

Suppose, for the moment, that we bite bullet and simply identify presumably higher-order mental properties with their lower-order disjunctive associates. If the latter count as ‘physical’ properties, then we would be acceding to the weak reductionist thesis that mental properties are physical properties. But so what? We can still maintain that mental properties are distinct from any of the disjunct properties – we can deny, for example, that mental properties are identical with neurological properties – and that’s the reductionist thesis that Kim is pushing. (Antony, 2003, p. 7; italics original)

The work went like this. Clapp (2001) defined disjunctive properties as the properties that are referred to by predicates that are formed by disjoining the names of legitimate properties. Block (1997), Clapp (2001), Antony (2003), and Witmer (2003) all argued in various ways that the identification of psychological properties with lower-level disjunctive

9 Clapp writes:

Let us define the notion of a properly disjunctive predicate: a disjunctive predicate ($\pi_1 \lor \pi_2 \lor \ldots \lor \pi_n$) is a properly disjunctive predicate if and only if (i) there is more than one disjunct $\pi_i$; (ii) each disjunct $\pi_i$ designates a legitimate property; and (iii) each $\pi_i$ designates a distinct property. The notion of a disjunctive property is now defined as follows: $P$ is a disjunctive property if and only if $P$ can be designated by a properly disjunctive predicate. (Clapp, 2001, p. 123; italics original)

It isn’t clear to me that we have a good hold on just what is being referred to by ‘$P$’ or any other ‘properly disjunctive predicate’ at the end of Clapp’s procedure; but Clapp’s procedure is much better than nothing.
properties is not much of a reduction and poses no problem about the projectibility of psychological properties, *so long as the physical disjuncts have, in some sense, relevant causal powers in common*. What I wish to focus on in this section is what goes into that ‘in some sense’.

To say that the properties that form a disjunctive property are in some sense causally similar is to say one of two things, either that (i) that the disjuncts have at least one causal power in common, or that (ii) the disjuncts have non-identical but similar causal powers. Clapp (2001) defends a version of (i). Block (1997) seems to defend a version of (ii). Let us take these in turn.

Clapp, after reviewing arguments against the (metaphysical, not just scientific) legitimacy of disjunctive properties, arrives at the conclusion that disjunctive properties are metaphysically acceptable provided they ‘overlap on a property’ (Clapp, 2001, p. 129). To say that they overlap on a property is to say the following:

If \((X_1 X_2 \ldots X_n)\) designates a legitimate property, there must be some nonempty set of causal powers \(d'\) such that (a) every possible thing that satisfies \((X_1 X_2 \ldots X_n)\) possesses every causal power in \(d'\), and (b) every possible thing that possesses every causal power in \(d'\) satisfies \((X_1 X_2 \ldots X_n)\). (Clapp, 2001, p. 129)

Which is to say that every property in the set \(\{X_1, X_2, \ldots, X_n\}\) bestows all of the causal powers in \(d'\), from which it follows that every property in the set \(\{X_1, X_2, \ldots, X_n\}\) has at least one causal power in common. Clapp writes that this ‘leads naturally to the following definition of realization:’

\[
P \text{ realizes } Q \text{ if and only if (def.), where } p \text{ and } q \text{ are the sets of causal powers constituting } P \text{ and } Q, q \subset p.\] (Clapp, 2001, p. 129)
This sort of subset account of realization does seem to get around the problem presented by Kim for the legitimacy of multiply realized psychological properties. But it comes at the cost of taking the other horn of Kim’s dilemma -- recall, as Block put it, if the argument is allowed to take place on Kim’s terms then Kim wins either way. If the disjuncts are causally heterogenous then so is the realized property; if, as Clapp and other non-reductivist disjunctivists argue, the disjunction is causally homogenous, then nothing blocks the reduction.

The set of causal powers $d'$ is itself individuative of a property -- the property upon which the members of the set $(X_1 x \lor X_2 x \lor \ldots X_n x)$ overlap; but then, plausibly, the realized property saved by Clapp’s method just is the property $D$ individuated by the set of causal powers $d'$. But the problem is that this property, $D$, does not look like the sort of property with which a non-reductivist would want to identify mental properties. $D$ is at the same level of nature as are the base properties that are members of the set \(\{X_1, X_2, \ldots, X_n\}\) (which is to say it is instantiated in the same objects) and $D$ is not a higher-order property with respect to the members of the set. Indeed, given that membership in the set \(\{X_1, X_2, \ldots, X_n\}\) is determined by having the set causal powers $d'$, $D$ is itself in the set \(\{X_1, X_2, \ldots, X_n\}\). $D$ is, so to speak, the only member of this set that matters; it will be had by anything that has any other member of the set, and having it is enough for having the realized property. Clapp notes (Clapp, 2001, pp. 132-136) that his account of realized properties results in a pyrrhic victory at best for non-reductive physicalism. I

---

10 Clapp’s account of realization is very much like Sydney Shoemaker’s subset account of realization as presented in (Shoemaker, 2001, 2007). I think that Shoemaker’s account of realization is subject to the same criticism that I will present for Clapp’s account.

11 While it is true that not every collection of causal powers is individuative of a property, in this case the property that is individuated is already known to be real: it is the realized property.

12 Thus, Antony’s claim, quoted above, that even on Kim’s on assumptions the realized property is distinct from any one of its realizers, is thrown into jeopardy. The realized property does not seem to be distinct from $D$.
add that it results in a total defeat of functionalism. This looks like a classical type-identity theory of mental states.

I turn now to Block’s adoption of option (ii). Block (1997) responds to Kim’s Disjunction Argument by trying to show that jade is a bad comparison case for psychological properties, not for reasons like those presented by Fodor, but because the disjunction of jadeite and nephrite is unprojectible while it is likely that the disjunctions of realizers of psychological properties are projectible. Unlike jadeite and nephrite, Block writes, it may turn out that realizers of psychological properties are subject to ‘impressive constraints’. Block makes this argument in terms of his ‘Disney Principle’.

One way of leading into my disagreement with Kim is to note that these special science kinds are typically not nomically coextensive with completely heterogeneous disjunctions of physico-chemical properties.

In Walt Disney movies, teacups think and talk, but in the real world, anything that can do those things needs more structure than a teacup. We might call this the Disney Principle: that laws of nature impose constraints on ways of making something that satisfies a certain description. There may be many ways of making such a thing, but not just any old structure will do. It is easy to be mesmerized by the vast variety of different possible realizations of a simple computational structure, say that of an and gate, which can be made of cats, mice and cheese (Block, 1995) as well as mechanical or electronic components. But the vast variety might be cut down to very few when the function involved is mental, like thinking, for example, and even when there are many realizations, laws of nature may impose impressive constraints. (Block, 1997, p. 120)

Block goes on to note that such factors as the laws of evolution and the limited numbers of physically available ways of making a thinking organism come into play as constraints, so that in general ‘there are reasons to expect less than total heterogeneity at both the design and
realization levels’ (p. 122) of psychological base properties.\textsuperscript{13} Block writes that this may be enough to meet the criterion for scientific respectability – projectibility -- that multiply realized properties must meet, though he is not certain. ‘How strong are the constraints imposed by the Disney Principle? We don’t know. And not knowing, we don’t know how right or how wrong Kim’s picture of science is’ (p. 124).

I suspect that what one makes of the choice between (i) and (ii) will really depend upon what one makes of projectibility. If mere similarity of causal powers is enough to guarantee projectibility, then Block’s method may be preferable -- though it seems one would have \textit{somehow} to state in advance of experiment what would count as sufficient similarity in experimental results, and to do so without appeal to identity of properties. If one suspects that this could not be done, that projectibility would have to be cashed-out ultimately in terms of identity of causal powers displayed in experimental results, then one will side with Clapp.

It seems to me that defenders of non-reductive physicalism who employ A-type role properties in their responses to Kim’s Disjunction Argument face a dilemma. If a philosopher wants to hold both that psychological properties are identical to A-type role properties and that A-type role properties are not wildly disjunctive, then the philosopher must adopt something like either Clapp’s or Block’s take on the nature of legitimate disjunctive properties. If one takes Clapp’s approach, one must deal with the apparent consequence that one has abandoned functionalism in favor of the identity theory -- though something like functionalism may survive as the method by which one arrives at the identities in question. If one takes Block’s approach,

\textsuperscript{13} Although I should add that Block qualifies the claim that evolution places constraints on psychological realizers by saying it doesn’t matter whether psychological regularities are adaptations or spandrels, since the ‘channels’ that evolution must move through, as dictated by the Disney Principle, will be the same either way. So what is really doing the work in constraining psychological realizers is the Disney Principle, not so much evolution as such. (Block, 1997, pp. 122-123)
one will have to find a way to cash-out projectibility by appeal to similarity, not identity, of causal powers. And in either case, one is essentially making the empirical bet that special science properties in general and mental properties in particular have disjuncts that are causally homogenous. Either way, one is making exactly the bet that Putnam said one was likely to lose, in his paper (Putnam, 1975) that introduced and motivated functionalism in the first place.

§6

In this chapter I examined an under-appreciated ambiguity in the common conception of role properties. I showed that this ambiguity reveals that there are two kinds of role-property that have different metaphysical characteristics. I argued that identifying special science properties in general and psychological properties in particular with either of these two types of role property requires abandoning a key aspect of traditional functionalism. If one identifies psychological properties with B-type role properties, and one wishes to avoid epiphenomenalism, then one will have to deny the principle that Kim calls the Causal Inheritance Principle. On the other hand, if one identifies psychological properties with A-type role properties, one will have to abandon robust multiple realizability and, quite possibly, embrace a form of type-identity theory.\footnote{The resulting picture both is and is not like Lewis’s version of the identity theory. It is like Lewis’s version in that the identities are arrived at via functional (causal role) analysis. It is not like Lewis’s in that mental properties are not taken to be species-specific.} Probably neither of these options will be acceptable to a philosopher who wishes to advocate a version of non-reductive physicalism; the first is non-reductive but in danger of not being physicalism, while the second is rather plainly reductive.

However, there is a much more basic concern for advocates of realization, here. Realization was supposed to be, all along, a relation between a base property and one unitary

125
realized property. I have argued that on the causal role-playing account of realization, realization turns out to be a relation between a base property and at least two distinct role-properties. The role-playing relation therefore does not look like the kind of relation that philosophers are after when they posit a realization relation. Further, the existence of this duality of role-properties seems to be the kind of thing guaranteed to result in philosophical squabbles for years to come, with some philosophers arguing for the identification of mental properties with A-type role properties, and some philosophers arguing for the identification of mental properties with B-type role properties. It may be that it would be better to abandon the role-playing account of realization in search of a tidier mind-body relation.
§1.1 The bulk of this dissertation has been devoted to a type of realization that I call “same-level realization.” Same-level realization itself has two subtypes, reductive and non-reductive. My primary goal has been to show that neither reductive nor non-reductive same-level realization may be explicated as causal role-playing. My motivations for doing so are (i) reductive and non-reductive physicalism in the philosophy of mind have become nearly co-extensive with the use of reductive and non-reductive causal-role realization in the philosophy of mind, (ii) the use of causal rolX-realization in the philosophy of mind cannot, I think, be made coherent, and therefore (iii) we need to begin anew the philosophical effort to understand just what reductive and non-reductive physicalism in the philosophy of mind are.

But there is another type of realization altogether, one that is not exclusively or even particularly used in the philosophy of mind. This other type of realization has become common in the philosophical literature, especially in the metaphysics of the special sciences. It can boast its own pedigree and its own proprietary uses. Philosophers who make use of this other type of realization do not claim that it can be explicated as causal role-playing and indeed may explicitly deny that it can. This other type of realization I will dub “cross-level realization”; the burden of this chapter will be to explore it.

It is probably easiest to explain what cross-level realization is by contrasting it with same-level realization. While same-level realization relates two of an object’s properties,
showing how (why, that) the object has one of them in virtue of its having the other one, cross-level realization relates a property of an object to many of the properties of that object’s proper parts. This is primarily why I have chose the name “cross-level realization” -- the relation “jumps,” as it were, from properties at one mereological level to the next, explanatorily relating them in a yet-to-be determined way.

Another difference between same and cross-level realization: as we just saw, same-level realization is a one-one relation; it relates a single realizer property to a single realized property. Only one realizer property at a time may “play the causal role of” a realized property -- the standard way of explicating just what same-level realization is. On the other hand, cross-level realization is, on the usual sort of account, a many-one relation -- it purports to explain how several of the properties and relations of an object’s several parts may work together to “realize” a single property of the object that the parts compose.\(^1\)

Another difference: when a property is same-levelly realized, that property is generally believed by philosophers to have, on that specific occasion of realization, the same causal powers as the property that realizes it on that occasion. To put that another way, if an object’s ability to enter into certain sorts of causal interaction is explained by its possession of a same-levelly realized property, then that object’s ability to enter into those causal interactions is also explained by its possession of the property that realizes the realized property on that occasion. This, indeed, is part of what is meant by saying that the realizer property “plays the causal role” of the realized property on a given occasion of same-level realization. On the other hand, in cases of

\(^1\) Sydney Shoemaker offers an account of cross-level property realization that he insists is one-one. Rather than relate numerous properties of an object’s parts to a property of the object, Shoemaker’s “microrealization” relates a (single) complex state-of-affairs that is itself made up of instantiations of properties in an object’s parts to a realized property of the object. We will examine Shoemaker’s account of cross-level realization later in this chapter, and see whether this distinction really makes a difference. (Shoemaker, 2007, chapter 3 passim)
cross-level realization, the causal powers that an object possesses due its having a realized
property are not also explained by its possession of the realizing property -- for the simple reason
that in cases of cross-level realization the object that has the realized property does not also have
a realizing property; rather, it is the object’s parts have the realizing properties.

But it is almost a secondary concern what object possesses the realizer properties in cases
of cross-level realization, since the causal powers of the realized property differ from the causal
powers of the realizer properties. To put that differently, as cross-level realization conceives of
its relata, the realized property figures in explanations of different sorts of causal phenemena
than could be explained by the mere possession of the realizer properties, either individually or
in sum. A cross-levelly realized property is thus qualitatively different than any of its realizers,
and it is qualitatively different than than the simple sum or conjunction of the causal powers of
those realizers.

§1.2 Here are the three characteristic differences between the standard accounts of same- and
cross-level realization that we just discussed, arranged in a table:

<table>
<thead>
<tr>
<th></th>
<th>Same-Level Realization</th>
<th>Cross-Level Realization</th>
</tr>
</thead>
<tbody>
<tr>
<td>What has the realizer property(ies)?</td>
<td>the object that has the realized property</td>
<td>parts of the object that has the realized property</td>
</tr>
<tr>
<td>Is the relation one-one or many-one?</td>
<td>one - one</td>
<td>many - one</td>
</tr>
<tr>
<td>does the realized property</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>have the same causal powers as its realizer(s)?</td>
<td>(all of them or a proper subset of them)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1
As a sort of extended aside, let me take two paragraphs to stress that the information presented on the right-hand column of Table 1 is a simplification; there is less agreement about the nature of cross-level realization among its advocates than the column suggests. Here are two examples. First, some advocates of cross-level realization believe that a realized property’s realizers may extend beyond the parts of the object that has the realizer, into the broader world, while some advocates do not. So it is not strictly true that all advocates of cross-level realization say that what has the realizers are (just) the parts of the object that has the realized property, as Table 1 suggests.

Second, while Table 1’s characterization of same-level realization works for both non-reductive and reductive versions of same-level realization, the characterization of cross-level realization applies only to non-reductive cross-level realization. This can be established easily. Recall that reductive realization identifies realized properties with their realizers. This is possible only if, for one thing, both the realizer and realized properties are had by the same object -- but that condition is violated by Table 1’s description of cross-level realization. So cross-level realization -- at least the sort described in Table 1 -- has no reductive subtype. This is forgivable because nearly all advocates of cross-level realization take it to be a non-reductive relation. But it should be pointed out that there is at least one philosopher of note who makes use

---

2 Carl Gillett confines cross-level realizers to the properties and relations of the parts of the object that has the realized property (2002, 2003); Sydney Shoemaker, and Carl Craver and Robert Wilson, allow some of a realized property’s realizers to be properties and relations of objects that are not parts of the object that has the realized property (Craver & Wilson, 2006; Shoemaker, 2007). There is a question here about how or whether to employ talk of “core” and “total realizers” when thinking about cross-level realization, and whether it is the core or total realizers that are supposed to extend beyond the object in the more permissive accounts of cross level realization. We will return to these questions later in this chapter.
of reductive cross-level realization, so it is a shortcoming of Table 1 that that spot in the logical space of options is not accounted for.

§1.3 Having taken note of some shortcomings of the table, let me now press what I think is the most important conclusion that we can draw from an examination of these characteristics of same and cross-level realization, and that is that same and cross-level realization are not in the same sort of explanatory business. This is a more fundamental distinction between same and cross-level realization than any we have discussed so far.

Suppose that we are interested in asking the following question: Why does this instance of this realized property convey these causal powers to the object that has it, rather than some other causal powers? Invoking same-level realization answers this question straightaway: the realized property conveys the causal powers that it does on this occasion because on this occasion it is being realized by this base realizer property, and in cases of same-level realization an instance of a realized property will have the same causal powers as its realizer. (Kim called this “the Causal Inheritance Principle”; we will say more about it later.)

But invoking cross-level realization leaves the italicized question unanswered. In cases of cross-level realization the causal powers of an instance of a realized property are not the same as the causal powers of any one of its realizers and they are not the same as a simple sum or conjunction of the causal powers of all of its realizers. This is important to see. It is not that cross-level realization is a many-one relation such that the realized property has all of the causal

---

3 The lone reductive cross-realizationist that I am aware of is John Heil, who presents a version of reductive cross-level realization in his book From an Ontological Point of View (Heil, 2003). Here, very briefly, is how it is supposed to work. Heil takes the lowest level of nature to be the only level of nature. Higher level objects and properties are “realized” by objects and properties at the lowest (and only real) level of nature not in the sense that higher-level objects and properties are “made real by” lowest-level objects and properties but in the sense that claims that refer to higher-level objects and properties are made true by lowest-level objects and properties. This is a sort of nominalist realization via indirect reference. I shall say no more about it, here.
powers of its many realizers, and therefore its causal powers cannot be explained by inheritance of the causal powers of any one of them. That’s not the idea. Rather, the idea is that the realized property is qualitatively different than any of its realizer properties or the conjunction of all of them. Thus, advocates of cross-level realization will say that the biological properties of a cell are realized by the properties of the chemical parts of the cell. But the causal powers enjoyed by a cell in virtue of its biological properties are not the same as the causal powers enjoyed by the molecules that make-up that cell in virtue of their chemical properties. The causal powers are not simply transferred from the molecules to the cell that they compose. Rather, the biological properties are distinct and different (as we saw, qualitatively different) from any or all of the chemical properties, though the cell has its properties in virtue of the chemical properties of its constituent molecules and the chemical properties result in the biological properties of the cell.

Therefore, and this is the point, if we want to know why the cell has these rather than those causal powers, it is not enough to say “because the chemical parts of the cell have the properties that they do, and those properties cross-levelly realize the biological properties of the cell.” There is still a further explanation to be given about why realization results in the cell having these properties with these causal powers rather than those properties with those causal powers. The explanation, to be sure, begins with the observation that the cell has the parts that it does, and that those parts have the chemical properties that they have. But the complete explanation for why these parts with these properties result in this whole with these properties will be provided by scientific research, and not merely by the observation that the biological properties of the whole are realized by the chemical properties of the parts.

---

4 We will discuss this “qualitative difference more when we get to our discussion of Carl Gillett. I borrow the expression “qualitative difference” from his new paper (forthcoming, p. 9).
Cross-level realization is therefore not explanatory in the same way that same-level realization is explanatory. When a property is same-levelly realized there is simply no further question to be asked about why this property with these causal powers has been realized rather than that property with those causal powers. But this is not to say that invocations of cross-level realization are not explanatory at all. To see the difference between the explanatory power of same- and cross-level realization more clearly, I would like to set up an analogy between these different kinds of realization and two different kinds of evidentiary relation that may come into play in a courtroom.

So I pause for a quick story.

§1.4

Smith is on trial for murder. The evidence against him is dire; it looks sufficient to get a conviction. However, at the last minute, the defense brings forward a new piece of evidence: a fingerprint of Smith’s found on a wall in an office building two hundred miles from the scene of the murder. The walls of the office building had been cleaned just an hour before the murder; no fingerprints had survived the cleaning. Furthermore, there is no chance that Smith had been in the office building at any time more than one hour after the cleaning. With this new piece of evidence before the court, the jury finds Smith not guilty of murder.

Reporters wait at the bottom of the courthouse steps to interview jury members as they come out. A reporter asks a jury member, “Why did you find Smith not guilty?” The juror responds, “The fingerprint exonerated him.” The reporter asks, “How so?” The juror looks at the reporter, somewhat annoyed, “The fingerprint stands in the exoneration relation to Smith.” The reporter is confused. “But how did the fingerprint exonerate him?” The juror rolls his eyes as he walks away. “I wasn’t being paid enough to ask such detailed questions.”

The exoneration relation is interesting. When we say that the fingerprint exonerates Smith – when we say, somewhat ostentatiously, that the fingerprint stands in the exoneration relation to Smith – we are saying that the fingerprint stands in some other relation to Smith that explains why Smith could not have been guilty of the crime of which he was accused. The
exoneration relation itself does not do that explanatory work. Thus, when the reporter is told that the fingerprint exonerates Smith, and is not told how, the reporter feels she has not been given an adequate explanation for just why Smith was found not guilty. To be sure, she was given a partial explanation – the reason Smith was found not guilty had something to do with the fingerprint – but being a good reporter she wanted to know more.

We can make a start at defining the exoneration relation like this:

(Exoneration) Smith stands in the exoneration relation X to the fingerprint iff there exists some other relation Q between Smith and the fingerprint such that $Q_{Smith, fingerprint}$ explains why Smith could not have committed the crime in question.

I will argue that cross-level realization is to X as same-level realization is to Q. In the remainder of this chapter I will occasionally refer to cross-level realization as X-realization (or as a form of X-realization) and to same-level realization as Q-realization (or as a form of Q-realization). Positing instances of Q-realization is more explanatory than positing instances of X-realization because Q-realizations explain why realized properties have the causal powers that they do while X-realizations do not. Positings of X-realizations are more like scientific promissory notes for such explanations.

When we get to exegeses of particular philosophers’ takes on cross-level realization, what we will see is that attempts to turn cross-level realization into a form of Q-realization require sacrificing one or more of the typical characteristics of cross-level realization seen in table 1.

§1.5 In the next few sections I will examine accounts of cross-level realization that have been put forward by three philosophers: Carl Gillett, Sydney Shoemaker, and Andrew Melnyk.⁵ I will

---

⁵ Shoemaker may seem out of place in this list. Shoemaker is best-known for his so-called “subset” view of realization, which is a version of same-level, not cross-level, realization. However he has recently put forward a version of cross-level realization, intended to supplement his account of same-level realization. He calls his version of cross-level realization “micorealization.” I will examine micorealization in this chapter. (Shoemaker, 2007)
not go into detailed discussions of their work; I will merely lay out their accounts of realization and discuss the ways that they meet or fail to meet the three characteristics of cross-level realization we examined above.

I will begin with the account that I believe best preserves the characteristics of “pure” cross-level realization, and that is Carl Gillett’s. From there we will see how other versions of cross-level realization sacrifice this or that aspect of pure cross-level realization in order to get a version of Q-realization.

§2

§2.1 In a series of papers over the past few years Carl Gillett has put forward and defended a version of cross-level realization that he dubs “Dimensioned Realization” (Aizawa & Gillett, 2009; Gillett, 2002, 2003). In addition, he has developed over that same time period a sophisticated readings of the history of metaphysical functionalism (2003, 2007, forthcoming). Gillett takes his own account of realization to grow out of his understanding of mistakes and confusions in the course of the development of functionalism. A study of Gillett’s views, and the use he makes of causal role playing, is therefore of obvious value.

I believe that I can perceive two phases of Gillett’s work on Dimensioned Realization in the 2000s. In the earlier papers (Gillett, 2002, 2003) he tries to get his version of realization, Dimensioned Realization, to drop out of a critique of what he calls “standard” or “flat” accounts of realization (what I call “same-level realization”). In later papers (Gillett, 2007, forthcoming) he goes the other way around, deriving a critique of the standard view of realization from its differences with his own view of realization, which he sees as being more in tune with the
sciences. What I will do, then, is discuss Gillett’s views in these two phases separately. In §2.2 I will discuss Gillett’s critique of the standard view and the way he took his Dimensioned Realization to drop out of that critique, and I will discuss why Dimensioned Realization on this early view ought to be seen as a version of X-realization. Then in §2.5 I will discuss Gillett’s later writings on Dimensioned Realization.

§2.2 Dimensioned realization is a cross-level version of the realization relation; more correctly, it is a version of realization that self-consciously folds same- and cross-level realizations into one definition of realization. We will see that Gillett takes this feature of dimensioned realization to grow out of his understanding of flaws in earlier versions of realization presented by others. We’ll get to that presently. First, here is an early version of his definition, dubbed “(D1):"

\[
(D1) \text{ Property/relation instance(s) } F_1 - F_n \text{ realize an instance of a property } G, \text{ in an individual } s, \text{ if and only if } s \text{ has powers that are individuative of an instance of } G \text{ in virtue of the powers contributed by the instances of } F_1 - F_n \text{ to } s \text{ or } s \text{'s constituent(s), but not vice versa. (Gillett, 2002, p. 322, 2003, p. 594) }
\]

I wish to point out two aspects of D1 right away. First, as mentioned, it combines same- and cross-level realization into one definition. It does this via the phrase “s or s ’s constituent (s).” If the base realizer property is had by s, the same individual that has the realized property G, then the realization relation is same-level. If the base realizer properties are had by s ’s constituents, then the realization relation is cross-level. At this point in the development of the definition, Gillett does not deny that the realized property must be a second-order property in the case of same-level realization; while it is clear enough that the realized property must be a first-order property in the case of cross-level realization. Therefore the same property could not be both same- and cross-levelly realized. Therefore, most importantly, the two kinds of realization
that are being combined are fundamentally different – their conjunction into one definition is rather ad hoc. This will change later on, when Gillett denies that a same-levelly realized property must be a second-order property.6

The second thing I wish to point out is that D1 is clearly an example of X-realization. D1 tells us that the realized property G is had by s “in virtue of” the property or properties F1 -Fn, but D1 does not purport to tell us how F1-Fn bring G about. Throughout Gillett’s writings, when it comes to explaining how any given F1-Fn bring about a given G, Gillett appeals to specific explanations provided by the sciences. To say that G is “realized” in the sense of D1, then, is to say that some scientifically kosher explanation for the presence of G exists; an explanation that appeals to properties F1-Fn of s or s’s constituents.

I should stress that the lack of a more specific explanation for the presence of G is not a flaw in the definition of dimensioned realization; it is rather a mark of D1 being a definition of a form of X-realization rather than Q-realization. It is worth noting, though, the absence of any appeal whatsoever to causal role-playing. Causal role-playing does feature in Gillett’s account of realization, but in different ways than it does in what he dubs the “Standard Picture” (2007, p. 23; forthcoming, p. 7) of functionalism. This aspect of Gillett’s view is obviously crucial for my purposes, and we will return to it.

First, though, having laid out Gillett’s definition of dimensioned realization, I wish to discuss how he takes his definition to be a correction of mistakes that he perceives in the Standard Picture of functionalism. In Gillett’s earlier papers he located two crucial flaws or

---

6 This may be unfair, as Gillett may have assumed that what he wrote explicitly later on was implicit in his (2002) and (2003), namely that it is never the case that a realized property must be a second-order property.
limitations in the Standard Picture. The first is a specific view explicitly held by many same-level realizationists; one that is codified by Jaegwon Kim’s Causal Inheritance Principle:

If a second-order property $F$ is realized on a given occasion by a first-order property $H$ (that is, if $F$ is instantiated on a given occasion in virtue of the fact that one of its realizers, $H$ is instantiated on that occasion), then the causal powers of this particular instance of $F$ are identical with (or are a subset of) the causal powers of $H$ (or of this instance of $H$). ((Kim, 1998, p. 54)\(^7\)

What the Causal Inheritance Principle codifies, and that Gillett thinks is a mistake, is the idea that the causal powers of a realized property must match the causal powers of the realizer. (Exactly why Gillett thinks this is a mistake, we will examine momentarily.) The second flaw or limitation in the Standard View is the idea that the base realizer property must be had by the same object that has the realized property. Gillett described these two aspects or flaws of the Standard view as follows:

(I) A property instance $X$ realizes a property instance $Y$ only if $X$ and $Y$ are instantiated in the same individual. (2002, p. 17, 2003, p. 593)

(II) A property instance $X$ realizes a property instance $Y$ only if the causal powers individuative of the instance of $Y$ match causal powers contributed by the instance of $X$ (and where $X$ may contribute powers not individuative of $Y$). (2002, p. 18, 2003, p. 593)

Of interest here is the direction of argument. Gillett took (I) to be a consequence of (II).

A central feature of the flat metaphysics [i.e. same-level realization] that bears emphasizing is the quite literal notion of “causal role playing” that it embodies.

\(^7\) Note that Kim’s (1998) version of the Causal Inheritance Principle differs from the (1993) version, which was the version discussed in the chapter of this dissertation on Kim, Fodor and Block) and which reads like this:

[The Causal Inheritance Principle] If mental property $M$ is realized at $t$ in virtue of physical realization base $P$, the causal powers of this instance of $M$ are identical with the causal powers of $P$. (Kim, 1993, p. 326)

The crucial difference is that the earlier formulation of the Principle required that the realized property have all of the causal powers of the base realizer property, while the newer version allows the base realizer property to have more causal powers than are had by the realized property. Shoemaker explicitly exploits this shift in Kim’s statement of the Principle to make room for his own subset account of same-level realization. (Shoemaker, 2007, pp. 16-17)
Under the flat view one property instance realizes another only if the realizing property contributes all the powers individuative of the realized property – the realized property instance thus literally plays the very causal role that individuates the realized property. *As a consequence,* the flat account makes no room for realization between properties in different individuals, for it appears that it is the very nature of instantiation that a property instance only contributes powers to the individual in which it is instantiated. If a realizer property instance matches the powers individuative of the realized property, the realizer must thus be instantiated in the same individual as the realized property. (Gillett, 2003, p. 594; italics added)

And again, arguing that (I) is a consequence of (II):

It is illuminating to consider a possible diagnosis of why the Standard View [same-level or “Flat” realization] has been attractive to philosophers, since this will be helpful in crafting a more adequate alternative. One explanation is that proponents of the Flat account have taken a quite literal view of a realizer X ‘playing the causal role’ of a realized instance Y, thus assuming X must contribute the very same powers as Y. Such literalism about role-playing leaves one committed to [the Causal Inheritance Principle]. But the very nature of the instantiation of a property in an individual is that this property instance contributes powers to just this individual. *Consequently,* given [the Causal Inheritance Principle’s] claim that realized and realizer properties coincide in the powers they contribute, one must conclude that realized and realizer instances are instantiated in the same individual. (Gillett, 2002, p. 321; italics added)

Given this order of reasoning, if one wanted to develop a new view of realization that involved the denial of both (I) and (II), one would think that the way to go would be first to find fault with (II), the Causal Inheritance Principle. Then, one would point out that since (I), the same-object restriction, follows from (II), the Principle, the denial of the principle allows one to deny the restriction, and therefore to explore cross-level realization. But that is not how Gillett argued. He went the other way around. He denied the same-object restriction and took this to be sufficient to deny the Causal Inheritance Principle.
In contrast [i.e. in contrast to the Standard Picture], the second metaphysical view – what I term the *dimensioned* account – allows that realized and realizer properties may be either instantiated in the same, or in different, individuals. This account *consequently* differs in a range of other respects from the flat view – not least in its notion of “causal role playing.” Under the dimensioned view, properties/relations $F_1$-$F_n$ of the constituents of an individual $s$ can play the causal role of a property $G$ of $s$ without $F_1$-$F_n$ contributing any of the causal powers of $G$. For in cases of realization involving different individuals, the realizer properties/relations instantiated in constituents do *not* contribute powers to the constituents that are individuative of the realized property $G$ instantiated in the constituted individual. Nonetheless $F_1$-$F_n$ play the causal role of $G$ in a wider sense, for the constituted individual has the powers individuative of $G$ in *virtue of* the distinct powers contributed by $F_1$-$F_n$ to the constituents. (Gillett, 2003, p. 594; italics added to "consequently")

Here, Gillett is claiming that a different understanding of causal role-playing *drops out* of the denial of the same-object restriction. Now, as a matter of logic, there is nothing wrong with this line of argument -- it is simply an instance of modus tollens. If (II) then (I); not (I); therefore not (II). But this will not do, here; it will not do simply to invoke the old adage that one philosopher’s modus ponens another philosopher’s modus tollens. Since Gillett took the Causal Inheritance Principle to be the primary aspect of the Standard Picture of functionalism, and therefore the central feature of the usual understanding of realization, he surely needs some independent reason to find flaw with (II). He can’t just say that (I), a consequence of (II), is unpalatable, and therefore (II), which by his own lights is the central feature of the standard definition of realization, has to go, so that we can make up a new definition of “realization” and be off and running.

Let me press this point. On Gillett’s understanding as presented in his earlier papers (Gillett, 2002, 2003), the Causal Inheritance Principle is practically the same thing as what he calls “literal” causal role playing. To say that property A of object $s$ literally plays the causal role of property B of object $s$ is just to say that the causal powers of $s$ (i.e. the things that $s$ can do)
that single s out as a possessor of property B are, all of them, things that s can do because it has property A. On the Standard Picture that is just what it means to say that A realizes B; that is why on the Standard Picture realization is explicated as causal role playing – the two relations are said to be the same thing. And that is why it makes no sense on the Standard Picture to say that a property of some other object than s literally plays the role of property B of s. So there is something very fishy indeed about saying, out of nowhere, that properties of s’s constituents may realize properties of s... and since this is the case, realization must not necessarily be causal role playing.

Or rather, this would be very fishy, if *dimensioned realization were supposed to be an alternative version of Q-realization*. But it isn’t. Rather, dimensioned realization is a version of X-realization. To say that property A dimensionally realizes property B is just to say that some scientifically kosher relationship exists between A and B that renders B assimilable into a physicalistic or at least naturalistic ontology. To put this in terms of my exoneration analogy, property A exonerates special science property B of the charge of ghostliness if and only if property A dimensionally realizes property B. But this should not be taken to be a bold claim about the limited number of ways that B may be exonerated -- it should rather (and merely) be taken to indicate the broad scope of the notion of dimensioned realization.

This is an important difference between dimensioned realization and causal role-playing. Dimensioned realization subsumes causal role-playing. If the A that exonerates B is had by the same object that has B itself, then the relation that does all the work will be causal role playing. Thus causal role-playing may explain why A exonerates B (and thus explain why A dimensionally realizes B). But it would be a mistake to think that causal role playing is itself a
form of X-realization. It isn’t; nor did Gillett take it to be – but it is rather what accounts for some instances of X-realization. To put that another way, on Gillett’s early view, dimensioned realization is the correct explication of X-realization, Q-realization is not a form of X-realization, but Q-realization is sometimes what accounts for instances of X-realization. As a point of secondary importance to Gillett’s project, same-level realization sometimes occurs, is (literal) causal role-playing, and is a type of Q-realization.

§2.3 Now let us move on to Gillett’s later work.

A more recent definition of Gillett’s version of cross-level realization is as follows:

(D2) Property/relation instance(s) F₁-Fₙ realize an instance of a property G, in an individual s under condition $, if and only if s has powers that are individuative of an instance of G in virtue of the powers contributed, under $, by F₁-Fₙ to s or s’s constituent(s), but not vice versa. (Gillett, forthcoming, p. 10; my boldface)

For present purposes I would like to ignore the boldfaced words “s or” as the simply complicate matters by bringing in Gillett’s laudable but for now beside-the-point attempt to define cross- and same-level realization all at once. Our discussion will be aided if we simply use this slightly simplified definition:

(D3) Property/relation instance(s) F₁-Fₙ realize an instance of a property G, in an individual s under condition $, if and only if s has powers that are individuative of an instance of G in virtue of the powers contributed, under $, by F₁-Fₙ to s’s constituent(s), but not vice versa.

As we will see, it is the aspect of dimensioned realization retained in D3 that most interests Gillett. (The inclusion of same-level realization in D2 seems to have been done mostly out of politeness to the “flatties.”) So now let us see how D3 works.

---

8 Gillett refers to same-level-only realization as “flat” realization (Gillett, 2003) and at least one proponent of same level realization has adopted the term (Shapiro, unpublished).
§2.3.1 The idea behind D3 is that a property of an object may be said to be cross-levelly realized provided the object has the property in virtue of properties had the object’s parts and the relations of the parts to one another. This may be read counterfactually: absent redundant realizers, the object, here $s$, would not have had property instance $G$ had $s$’s constituents not had property and relation instances $F_1$-$F_n$. The first thing to note is that on a sort of “high metaphysics” reading D3 appears to make every property instance of $s$ cross-levelly realized. Assuming a modest version of the thesis that composition is identity, $s$ would not exist at all if it did not have parts that stand in the relations to each other that they do. There is no property instance $G$ had by $s$ that would exist in the absence of $s$, and so there is no property instance $G$ had by $s$ that is not there in virtue of properties and relations had by $s$’s parts.\(^9\)

However, I think that this is an overly lawyerly way to understand D3. Gillett’s intent is not to prove that all of a composed object’s properties are cross-levelly realized, but rather to provide an abstract explanation for how the causally (scientifically) relevant properties of an object result from the causally (scientifically) relevant properties of the object’s parts. The goal, that is, is to provide a metaphysical framework for the special sciences, not to arrive at “interesting philosophical conclusions about properties.” What is really of interest, for present purposes, is whether this makes D3 a definition of a form of X-realization or Q-realization. If D3 does not define a form of Q-realization then D3 is not in competition with accounts of same-level realization that explicate realization with the causal role-playing relation. So let us attend more closely to the spirit as well as to the wording of Dimensioned C-Realization.

\(^9\) I am operating on the assumption that a property instance is an entity that constitutively involves an object and a property. Change either one, and one is left with a different property instance. This is certainly true of property instances on Sydney Shoemaker’s definition of property instances as states-of-affairs (Shoemaker, 2007, p. 3, f. 3), and Gillett explicitly derives his theory of properties from Shoemaker (Gillett, 2002, p. 317, f. 2; 2003, p. 593, f. 6). Lately, Gillett has been developing his own theory of property instances, but it will be something of a surprise if his theory does not sustain the requirement to which I am here appealing.
I believe that what is relevant is a significant distinction in D3 between the property instances had by s’s constituents and the property instance had by s. F1-Fn are said to “contribute” causal powers to s’s constituents, while G is said to be “individuated” by causal powers that s has. Thus we seem to have different sorts of explanation for, on the one hand, why s’s constituents have the causal powers that they have and, on the other hand, why s has the causal powers that it has. s’s constituents have the causal powers that they have because they have property instances F1-Fn and instances of the properties of which F1-Fn are instances “contribute” these powers to the objects that have them. For s’s constituents, the possession of property instances explains the possession of powers. When it comes to s, though, it appears that it works the other way around: the having of the powers explains the having of the property instances. s has G because s has powers “individuative” of G.

Why does s have powers individuative of G? Not, evidently, because s has G, but rather because s’s constituents have the causal powers that they do. But s’s parts have the causal powers that they do because those powers have been “contributed” to s’s constituents by F1-Fn. It follows that s has G because s’s constituents have F1-Fn.

Evidently the properties of which F1-Fn are instances are an altogether different kind of property than is G. Possession of the properties of which F1-Fn are instances explains the possession of certain causal powers; but it is the possession of certain (other) causal powers that explains the possession of properties like G. For clarity let me define some more terms:

\[
\begin{align*}
Z_1-Z_n &= \text{proper parts of } s \\
G &= \text{a property had by } s \\
H &= \text{the instance of } G \text{ had by } s \\
h &= \text{the causal powers had by } s \text{ that are individuative of } H \\
E_1-E_n &= \text{the properties had by } Z_1-Z_n \text{ of which } F_1-F_n \text{ are instances} \\
F_1-F_n &= \text{the instances of properties } E_1-E_n \text{ that are had by } Z_1-Z_n
\end{align*}
\]
\[ f_1-f_n = \text{causal powers had by } Z_1-Z_n -- \text{powers of the type that } F_1-F_n \text{ contribute to objects under conditions } \$

The idea is that we have an object \( s \) with some proper parts \( Z_1-Z_n \). \( Z_1-Z_n \) have properties \( E_1-E_n \). We need to give names to the specific instances of \( E_1-E_n \) that are had by \( Z_1-Z_n \), so we call the instances “\( F_1-F_n \).” \( Z_1-Z_n \) also have certain causal powers, \( f_1-f_n -- \text{the explanation for } Z_1-Z_n \text{'s possession of } f_1-f_n \text{ is that } F_1-F_n \text{ contribute } f_1-f_n \text{ to } Z_1-Z_n. \) More generally, in circumstances \( \$ \), instances of \( E_1-E_n \) contribute causal powers of the same type as \( f_1-f_n \) to the objects that have them. Right now, \( Z_1-Z_n \) both have \( E_1-E_n \) and are in \( \$ \) so \( Z_1-Z_n \) have \( f_1-f_n \). The question, now, is: how do we get from these facts about \( Z_1-Z_n \) to an explanation for the fact that the composed object \( s \) has \( H \), an instance of the property \( G \)? The answer is that \( s \) has certain causal powers \( h \); and \( s \) has \( h \) in virtue of \( Z_1-Z_n \)’s possession of \( f_1-f_n \). \( Z_1-Z_n \)’s possession of \( f_1-f_n \) non-causally results in \( s \)’s possession of \( h \) (Unpacking the “in virtue of” and the “non-causally results in” is a task for the sciences, on a case-by-case basis.) But \( h \) is individuative of \( G \) -- an instance of \( G \) comes into existence by there being an object with causal powers individuative of such an instance. The instance of \( G \) had by \( s \) is \( H \). Thus, and here is what I take to be the telling upshot of this, causal powers \( h \) explain instance \( H \) but instances \( F_1-F_n \) explain causal powers \( f_1-f_n \). At the lower level, the instances contribute and thereby explain the powers. At the higher level, the powers individuate and thus explain the instances.

It would be natural but wrong to say that, even so, \( H \) “contributes” \( h \) to \( s \) in just the same way that \( F_1-F_n \) “contribute” \( f_1-f_n \) to \( s \)’s constituents. To suggest that \( H \) “contributes” \( h \) to \( s \) while also keeping with everything that is said in the definition of Dimensioned C-Realization is to suggest that \( s \)’s possession of \( h \) has two separate explanations: (i) \( s \) has \( h \) because \( s \) has \( H \), (ii) \( s \) has \( h \) because \( s \)’s constituents have \( f_1-f_n \). But that cannot be the right idea; what Gillett clearly
wants is (ii), and there is no reason to insist on (ii), or on cross-level realization at all, for that matter, if one is willing to grant (i). But (and this again is the “telling upshot” I referred to earlier) this seems to make G and E₁-Eₙ very different kinds of property.

Here I think is how Gillett would respond. Gillett considers as a forerunner to his Dimensioned Realization a view of realization developed by William Lycan in his 1987 book *Consciousness* (Lycan, 1987) (Gillett, 2003, p. 321, f. 6). There, Lycan wrote that whether a property is counted as a realized property or a realizer property may be relative to the goals of an investigation (Lycan, 1987, pp. 37-39). Gillett writes that a one-level analysis of the functions of the parts of a mechanism will result in one way of understanding the origins of the causal powers of the parts. On the other hand, a cross-level analysis of a part will result in a different understanding of the causal powers’ origins. In practice, this means that Gillett is willing to say that the explanation for an object’s possession of causal powers is relative to the goals of an investigation. If we are doing an “intra-level” mechanistic analysis of a property, we may legitimately say that the property contributes causal powers to its objects. But if we are doing a cross-level analysis of that same property, we should say that the property is individuated by its causal powers; causal powers that the object with the property has in virtue of the causal powers of the object’s parts.

§2.3.2 It will be helpful to have a look at the general metaphysical picture of the world that seems to arise from these considerations. Evidently all properties may be seen as either contributing causal powers to objects or as being individuated by those causal powers while the actual explanation for the presence of those causal powers coming from “down below.” This obviously would not be true for the properties of objects at the bottom level, however; that, it
would seem, is where the buck stops. So let us assume that at the bottom level of nature there are basic objects and fundamental properties. At this level, it is simply true, a brute fact of nature, that there exist such-and-so fundamental properties that contribute such-and-so fundamental conditional causal powers to the basic objects that have them. The explanation for the presence of the fundamental causal powers in the basic objects is simply and literally that the fundamental causal powers are “contributed” in some unanalyzable sense to the basic objects by the fundamental properties. It is also true, but merely incidental (a mere epistemic point) that the fundamental properties may be individuated by their causal powers; but the fundamental property instances do not exist in the objects in which they exist due to individuation; the properties are simply had, brutely, by the basic objects that have them,

The fundamental objects combine to form composed objects at “the next level of nature up.” These composed objects have causal powers in virtue of the fundamental properties and relations of the basic objects that compose them. These “next-level” causal powers may then be used to individuate properties. A “property instance” at this second level of nature is simply a set of causal powers had by a specified composed object. The properties have a nominal feel to them; what more fundamentally exists at the second level of nature are the causal powers had by the composed objects, powers cooked-up by the combination of fundamental powers at level one. At level three there are objects composed of objects that exist at level two. And so on.

It may at this point be an interesting fact that the way in which the causal powers of the objects at level two combine or mix to create the causal powers at level three may be better understood by grouping the level two powers into sets that we call “properties and relations” and that may give an independent certification to the existence of the level two properties and
relations, so that the level two properties and relations are not merely nominal collections of
powers. Nevertheless, while it may be convenient to say from the point of a view of an analysis
of the properties of objects at level three that the causal powers of objects at level two are
“contributed” to them by their properties and relations, it is not literally true. What is literally
true is that the objects at level two have the causal powers that they have in virtue of the causal
powers of their constituents, in just the same way that the causal powers of the objects at level
three are there not in virtue of level three properties and relations, but in virtue of level two
causal powers.

So what we have is a world made of basic objects and fundamental properties and
relations. Perhaps we should regard the instances of the properties and relations as primitive,
too. The basic objects have such and so properties and relations primitively. The fundamental
properties and relations contribute causal powers to the basic objects, where this “contribution”
relation is also primitive. From there, everything else is composition and causal powers. There
are no “higher level properties” except quasi-nominalistically.

I do not know if Gillett would endorse this picture of the world, but I believe it is the
picture of the world that drops out of his definition of cross-level realization. Though this sort
of realization can be described as a relation between properties, it would be more direct to
describe it as a relation between causal powers. I develop this point in §2.5.

§2.3.3 Since 2002 Gillett has used two primary examples of cross-level realization to both derive
and illustrate aspects of dimensioned realization. In earlier papers his favored example was the
property of hardness possessed by a diamond, realized by the properties and relations of the

\[10\] It is also strikingly similar to the picture of the world that Kim attributes to Shoemaker in Kim’s discussion of
Shoemaker’s book Physical Realization (Kim, 2010). I take it that this is not a coincidence, since Gillett adopts
Shoemaker’s causal theory of properties and a great deal of the world-picture drops out of that theory of properties.
diamond’s constituent atoms (Gillett, 2002, 2003). (He used this example again in a later paper co-authored with Ken Aizawa (Aizawa & Gillett, 2009)). I would like here, though, to focus on the other of Gillett’s favored examples, one that he began employing in 2007.

Gillett discusses a part of a neuron called a potassium ion channel, and its property of being a voltage sensitive gate. Considered as a part of mechanism that operates at one level of nature, the potassium ion channel has the power to open in response to changes in surrounding cells, allowing potassium ions to pass through. The channel has this power (on a one-level analysis) because the power is contributed to the channel by the channel’s property of being a voltage sensitive ion gate.

However, we can also delve more deeply into the workings of the ion channel, and seek to explain how the powers possessed by the parts of the ion channel non-causally result in the channel’s power to open in response to a change in the charge of surrounding cells. Considered this way, the channel’s power to open is explained by the powers of its parts, not by the channel’s possession of the property of being a voltage sensitive gate.

Such inter-level mechanistic explanations allow us to understand the properties and other entities posited in intra-level explanations by illuminating lower level entities that compose them. For example, in the case at hand, Roderick MacKinnon won the Nobel prize for his work establishing a compelling inter-level mechanistic explanation of how such ion channels open by illuminating the chemical and spatial properties/relations of the complex protein molecules that are ‘sub-units’, i.e. parts, of these channels. Basically, as Figure 1 illustrates, when the charge in surrounding cells changes, then the backward-looking powers, to change relative spatial position, of each of the sub-units are manifested together and the sub-units all swivel to adopt new spatial relations in relation to each other. These lower level mechanisms together implement the higher level process of the ion channel opening. And the many properties and relations of the sub-units, such as their alignment and chemical properties, together realize the ion channel’s property of being a voltage sensitive gate – basically, the powers contributed by the lower level properties together non-causally result in the qualitatively different powers of the realized property. (Gillett, forthcoming, p. 8)
Figure 1. A diagram of the lower level mechanisms at bottom, involving the protein sub-units, that implement the process of the potassium ion channel opening in response to a change in charge of nearby cells, outlined in the top of the figure. (Figure and caption from Gillett (forthcoming).)

From this example Gillett derives three characteristics of dimensioned realization.

First, dimensioned realization is a transitive, asymmetric, non-causal ontological determination relation. It is non-causal in that it is “synchronous, occurs between entities that are not wholly distinct, and do not involve the mediation of force and/or transfer of energy” (Gillett, forthcoming, p. 9).

Second, the properties that are related by dimensioned realization are qualitatively different: “Thus the properties and relations of the protein sub-units contribute no common
powers with the property of the ion channel -- the protein sub-units have powers such as changing their relative spatial positions, under certain conditions, but not the powers of opening or allowing speedy passage to potassium ions” (p. 8).

Third, dimensioned realization is many-one, and this many-one character explains the qualitative difference of the realizer and realized properties.

Thus although the properties/relations of the sub-units share no common powers with the property of being a voltage sensitive gate, nonetheless together the contributions of powers by the properties/relations of the sub-units can non-causally result in the powers individuative of this very different property. (p. 8)

§2.3.4 Using these three features, derived from the example, Gillett constructs the definition of dimensioned realization that was quoted in §2.3.1. However, we now have a puzzle. Dimensioned realization purports to be an explanation for why composed objects have the property instances that they do. But in fact the explanation for the presence of the higher-level property instances that is given to us here does not seem to be realization at all -- rather, it is individuation. The channel has the property of being an voltage sensitive gate because it has causal powers individuative of the property of being a voltage sensitive gate.

The relation that holds between the realized property instance (an instance of the property of being a voltage sensitive gate) and its realizer property instances (instances of chemical and alignment properties had by the protein subunits) appears to be indirect and non-explanatory. What does the explanatory heavy lifting is the relation between the powers of the subunits and the powers of the channel. We are told by Gillett that the powers of the subunits “non-causally result in” the powers of the whole. That is the relation of interest.

The discovery that wins the Nobel prize -- the discovery for which dimensioned realization purports to provide a metaphysical scaffolding -- is the discovery that illuminates the

151
relation between the powers, not the indirect relation between the properties. They way in which the powers of the subunits “work together” to result in the powers of the channel is what demands scientific investigation. In the absence of a clear understanding of the way in which the powers of the subunits work together, there is no explanation to be had for the presence of the realized property, other than that it is individuated by powers possessed (inexplicably) by the whole. That is what makes Dimensioned C-realization a type of X-realization rather than a type of Q-realization.

The explanation for the presence of the powers of the whole is that the powers had by the whole’s parts “non-causally result in” the powers of the whole. But now it appears that the three characteristics that Gillett attributed to the relation that holds between the properties of the parts of the channel and the properties of the channel as a whole are actually characteristics of the relation that holds between the powers had by the parts of the channel and the powers had by the whole. To revert to my previous terminology, powers f₁-fₙ non-causally result in h, and the relation between f₁-fₙ and h is transitive and asymmetric. The powers are qualitatively different, and the relation is many-one. Since the protein sub-units have f₁-fₙ because f₁-fₙ are contributed to the subunits by F₁-Fₙ, and since h is is individuative of H, then in some indirect sense, to be sure, F₁-Fₙ “non-causally result in” H, but it is hard to see why this indirect relation between property instances deserves the name “realization” more than does the relation between f₁-fₙ and h. It is not literally true that F₁-Fₙ make H real; but it is literally true that f₁-fₙ make h real. (Though in exactly what sense f₁-fₙ make h real is a topic to which we will return.)

Let us contrast this picture of cross-level realization with the standard picture of same-level realization. In same-level realization the relation between properties is much more direct:
a realizer property contributes powers to an object that are individuative of a realized property. That is what it means to say one property “plays the causal role” of the other and that is why the realized property must directly inherit the causal powers of its realizer if it is to have any causal powers at all. A realized property is merely individuated . . . made real, by the causal powers of some other property. So the realizer property “makes real” the realized property in a straightforward way. (Or anyway it would, if the causal role-playing explication of same-level realization ultimately worked, which I have argued it does not.)

Once it is seen that the protein subunits have the power to change their alignments relative to one another in response to a change in the charge of surrounding cells, it becomes clear why the potassium ion channel that is composed by the subunits has the power to open in response to a change in the charge of the surrounding cells. Furthermore, it is clear why the channel has the power to open in response to the charge of the surrounding cells, rather than some other power. Therefore it is clear why the channel has the property of being a voltage sensitive gate, rather than some other property.

But to say this is not yet to say just what is doing the explanatory work. The question to ask is: is there some cross-level analog to the causal role-playing relation that is doing the explanatory heavy lifting here? Not, evidently, the cross-level realization relation itself. Cross-level realization does not explain how the many combine to result in the one -- it just affirms that they do. That is why cross-level realization, when it is taken to be a many-one relation, is like exoneration.

---

A flattie would say that the channel changes its shape in response to changes in charge, and that is what realizes the channel’s property of being a voltage sensitive gate. The flattie would add that there is no other way to realize the property of being a voltage sensitive gate. True, there are many ways to make an entity that changes its shape in the way the potassium ion channel does (constructing it out of rope-shaped parts that untwist in response to a change in charge is only one of them) but that is not to say that there is more than one way to make a voltage sensitive gate: one must make an entity that changes its shape in response to changes in ambient charge. See Polger (2008).
In the next section, then, we see that Sydney Shoemaker attempts to develop a form of cross-level realization that is one-one rather than many-one. This attempt is understandable, given what we have said, since it appears to be necessary for a form of realization to be one-one in order for it to be a form of Q-realization -- that is, for it to be as robustly explanatory as same-level realization. But what we will see in §4 is that attempts to make cross-level realization a one-one relation end up collapsing cross-level realization into same-level realization.

§4

§4.1 Shoemaker begins the chapter on cross-level realization in his recent book on realization by distancing his view of cross-level realization from Gillett’s. Shoemaker writes:

Such views [as Gillett’s] are right in holding that we need an account of realization that gives a role to the properties of micro-entities and other parts of macro-objects, and that we do not get this in the sort of [same-level] account presented so far. But the cure for this is not to count properties of parts of macroscopic objects as realizers of properties of the macroscopic objects. The instantiation of a realizer of a property should be sufficient for the instantiation of that property, and no property of a micro-entity that is a part of a thing is such that its instantiation is sufficient for the instantiation of any of the properties of that thing. (Shoemaker, 2007, p. 32)

Shoemaker is fond of the mantra *realizers are sufficient for what they realize*. In my view, relying upon this mantra too much will encourage one to conflate two distinct notions of realizers: realizers as causal role-players and realizers as strong subveners. Only strong subveners need be sufficient for what they realize. (Another way to make this point is to say that Shoemaker’s mantra encourages a conflation of core and total realizers, but only total realizers need to be sufficient for what they realize.) Nonetheless Shoemaker’s remarks do present an interesting initial question for Gillett’s view: Gillett presumably thinks of the individual
properties and relations that are realizers of macro-properties as *neither* causal role-players *nor* as strong subveners. So in what sense for Gillett are the individual micro-properties and relations that together realize a macro-property, realizers?

I believe that Gillett would say: in no sense at all. Gillett’s view is explicitly that the micro-properties and relations that realize a macro-property do so as a group, together.\textsuperscript{12} Cross-level realization for Gillett is a group activity. More to the point: Gillett’s dimensioned realization is a many-one relation.

Shoemaker writes (continuing directly):

> What is true is that the instantiation of a property of a micro-entity can be part of a state of affairs that is sufficient for the instantiation of a property of a macroscopic entity. What we have here is the realization of a property instantiation, not by another property instantiation, but by a microphysical state of affairs involving the instantiation of micro-properties in micro-entities. Such a state of affairs “makes real,” constitutes, the occurrence of a property instance. (Shoemaker, 2007, p. 32)

Shoemaker calls this kind of realization “microrealization.” It is evidently a one-one relation between a state-of-affairs and a realized property. Thus, Shoemaker’s version of cross-level realization dispenses with one of the characteristic features of cross-level realization that I listed in Table 1. Questions arise.

The first and most important question is this: on Shoemaker’s view does the realized property have the same causal powers as the microphysical state of affairs that realizes it (or a proper subset of them), or are the causal powers of the realized wholly different? To put that differently, are the two entities, the microphysical state of affairs and the realized macro-property instance, causally *qualitatively* different from one another?

\textsuperscript{12} See for example Gillett (forthcoming, p. 9).
The answer to this appears to be that the causal powers of the realizing state-of-affairs is the same as the causal powers of the realized property instance. They are qualitatively the same. Although Shoemaker’s writing in the chapter on microrealization is dense, and not as clear as it could be, he says two key things that I believe affirm this interpretation; two things that clarify what it is for a microphysical state-of-affairs to realize a property in a composed object, and also what the relationship is between the causal powers of microphysical state-of-affairs and the causal powers of the realized object. First, Shoemaker writes that the causal profile (the collection of causal powers) of the microphysical state-of-affairs must be *isomorphic* to the causal profile of the realized property.

Such an isomorphism will pair types of microphysical states of affairs with properties; and a particular microphysical state of affairs will realize a particular property just in case the state of affairs belongs to a type that is paired with that property. When this is so the existence of the microphysical state of affairs will realize the property instance. (Shoemaker, 2007, p. 41)

So one way of characterizing the relationship between the causal profile of the realizing microphysical state-of-affairs and the causal profile of the realized property instance is that they are isomorphic.\(^{13}\) But isomorphism is cheap, and isomorphism doesn’t explain *why* the property instance must occur when the realizing microphysical state-of-affairs occurs. Let us look further. Shortly after the above quote, Shoemaker writes:

Corresponding to a property P there will be a microphysical state of affairs type that is a disjunction, perhaps infinite, of more specific types of microphysical states of affairs. What unites these more specific state of affairs types is the fact that the forward-looking causal features in their causal profiles match those of

\(^{13}\) At this point, Andrew Melnyk decides that Shoemaker’s microrealization does not result in a robustly physicalist view of the world, on the grounds that the mere isomorphism of microphysical to functional properties does not show that the functional properties are not emergent. I share the worry but I think that Shoemaker must have in mind a substantive relationship between the microphysical properties and the functional properties -- a substantive relation that blocks emergence. See Melnyk (2006, pp. 151-152). I will return to this point in §4.4.
property P . . . Any state of affairs that is of one of these types will realize property P. (Shoemaker, 2007, p. 42)\textsuperscript{14}

So according to Shoemaker, a microphysical state-of-affairs realizes a property instance in a composed macro-object just in case the causal profile of the microphysical state of affairs is isomorphic to and matches the causal profile of the property instance of the macro-object. That is as much detail as Shoemaker gives us. The micro-parts of the composed macro-object have properties and relations. This collective having forms a (single) microphysical state-of-affairs. The causal profile of this state-of-affairs is isomorphic to and matches the causal profile of a property instance that is had by the macro-object that is composed by the micro-parts that participate in the micro-state-of-affairs.

I find that I cannot interpret this account except as saying that the causal profile of the microphysical state-of-affairs must be the same as the causal profile of the property instance of the macro-object. But if that is right then we can delete another of the distinguishing features of cross-level realization listed in Table 1 from Shoemaker’s account of cross-level realization. We have already deleted the many-one-ness of Gillett’s pure cross-level realization. It appears that we must also delete the qualitative difference in causal powers. It appears that we can’t keep the second after we have discarded the first.

We should now, I think, be wondering whether Shoemaker’s microrealization counts as cross-level realization at all. We could try applying the only remaining characteristic from Table 1, but it does not seem to apply to microrealization, one way or the other. According to the final characteristic, the realizer properties must be had by the proper parts of the object that has the realized property; but according to Shoemaker’s microrealization the property instances had by

\textsuperscript{14} The ellipses elide a phrase that adds backward-looking causal features to the formula. Shoemaker decided later that he could ignore those, so I will, too.
micro-parts are not cross-level realizers at all. Rather, complex states-of-affairs are, as described.

I think that a final consideration weighs in favor of my contention that Shoemaker’s cross-level realization collapses into same-level realization. Shoemaker holds that all properties that are cross-levelly realized are also same-levelly realized. Or, in his terms, all properties that are microrealized are also property-realized:

Every case of microphysical realization is also a case of property-realization. To every kind of microphysical state of affairs there will correspond a property, namely the property something has at that time just in case a state of affairs of that sort occurs in its career at that time -- so whenever an instance of a property is realized by a microphysical state of affairs of a given kind, it will be realized by an instance of the property corresponding to that kind of state of affairs.
(Shoemaker, 2007, p. 35)

Shoemaker calls the sort of property that corresponds to a type of microphysical state of affairs a “microphysical state of affairs embedding property” or “MSE” for short. Let us use an example to see how this is supposed to work. Let us say that this rock has the property of being smooth. The property instance of smoothness is microrealized by a microphysical state of affairs $msa$ (which means the rock itself is realized by $msa^{15}$) and it is realized by the MSE property that corresponds to $msa$. Plainly, since MSE property realizes smoothness in the normal same-level way, then via Shoemaker’s subset account of property realization smoothness must have for its causal profile a subset of the causal powers of MSE. But $msa$ also realizes smoothness, so the

---

15 Shoemaker writes:

It seems obvious that a microphysical state of affairs that minimally realizes the instantiation of a property at a time must realize the existence at that time of an object that is the subject of the property. I take realizers to be sufficient for what they realize [the mantra again -- DK], and plainly the instantiation of a property entails the existence of something in which it is instantiated. It also seems obvious that if a microphysical state of affairs realizes the existence of an object, some or all of the constituents of the state of affairs must be among those of which the object is composed. (Shoemaker, 2007, p. 36)
causal profile of \textit{msa} must be \textit{isomorphic to} and \textit{match} the causal profile of smoothness. I wrote above that I cannot make much sense of that unless it is allowed that the causal profile of smoothness is just the same as the causal profile of \textit{msa}, but perhaps it will do to say that the causal profile of smoothness is a subset of the causal profile of \textit{msa}. But now the causal profile of smoothness is a subset of both the causal profile of \textit{msa} and the causal profile of MSE. But given that \textit{msa} realizes the rock as well as MSE and smoothness, it seems that \textit{msa} is the rock, or at least, and I think this is the key formulation: \textit{the rock qua smooth}.

This may seem reductionist, since it reduces the rock and its properties to a microphysical state of affairs. Shoemaker will deny that his account is reductionist, on the grounds that smoothness could also be realized by non-physical microproperties in other possible worlds, but he will accede that aside from that it is reductionist. So I think the charge of reductionism is one that he would, with the mentioned caveat, accept. He replied to a similar charge from Jaegwon Kim in the 2009 APA Pacific session as follows:

\begin{quote}
At the end of his comments Jaegwon poses a dilemma: either my physicalism doesn’t count as physicalism, or it is a version of reductionist type physicalism. I certainly regard myself as a physicalist, and I think I accept the “starkly physicalist” picture Jaegwon finds suggested by some of what I say. I think that mental properties are physical in the same sense that automotive properties, architectural properties, computer properties, and botanical properties are physical. I would express this sense by saying that instances of all such properties are physically realized—they are both property-realized in accordance with my subset account, and realized in microphysical states of affairs in the way I spell out in Chapter Three. What matters for physicalism is that the actual world instances of these properties are physical, and it is sufficient for that that these instances are physically realized . . . As for whether my physicalism is reductive type physicalism, I think that it isn’t—while I hold that all actual world realizers of mental properties are physical, I am not committed to mental properties being necessarily such that they can only be realized physically for I am agnostic about whether there are dualist worlds in which mental properties have nonphysical realizers. For the same reason I am not a reductive type physicalist about automotive properties and botanical properties. (Shoemaker, 2010, pp. 125-126)
\end{quote}
§4.4 Andrew Melnyk, in a 2006 look at Shoemaker’s microrealization, criticizes it on the
grounds that it does not guarantee that instances of non-physical properties with physical
realizers are fully physically explained. For that reason, Shoemaker’s microrealization fails to
ground a true version of physicalism. Melnyk characterizes Shoemaker’s microrealization as
follows:

Microphysical state of affairs m realizes non-physical property-instantiation p iff
(1) the microparticles involved in m are among the microparticles that make up
the subject of property-instantiation p; and
(2) m belongs to a type of microphysical state of affairs M and p is an
instantiation of a property P such that M is paired with P in accordance with a
one-one mapping between a series of types of microphysical states of affairs and
the properties that a persisting thing has over time, where, relative to this
mapping, there is an isomorphism between the causal profiles of the types of
microphysical states of affairs and the causal profiles of the properties. (Melnyk,
2006, p. 151)

The problem, writes Melnyk, is that even if we read condition 1 as saying that the object
that has P is completely decomposable into the microparticles involved in m, it is still not
guaranteed that the macro-object’s having of P is robustly explainable by the microphysical state
of affairs m. The instance of P would only count as fully explainable in purely physical terms if
some “substantive metaphysical relationship between M and P” were present (Melnyk, 2006, p.
151). But the only relationship that is asserted to be present in (Melynk’s reconstruction of)
Shoemaker’s formulation of realization is isomorphism, a one-one mapping of causal profiles.
The reason for the mapping is not claimed to be explainable in terms of the realizer. Therefore,
Melnyk argues, Shoemaker’s version of realization does not count as a version of physicalism.
What would be needed in order to get a satisfying physicalism, according to Melnyk, is a “substantive metaphysical relationship” between M and P (ibid.). I end my discussion of Shoemaker’s microrealization with a discussion of Melnyk’s critique because Melnyk’s own version of realization can be understood as designed to meet exactly this objection. Melnyk does posit a “substantive metaphysical relationship” between physical realizers and what they realize. But, as I will show, the relationship the Melnyk posits is *brute necessitation*. This being so, the version of cross-level realization that Melnyk develops is not explanatory.

Thus, we began with a “pure” version of realization, Gillett’s, that met all three of the conditions of cross-level realization listed in Table 1. We saw that pure versions of cross-level realization count as versions of X-realization, not Q-realization. Next, we looked at Shoemaker’s attempt to develop a version of cross-level realization that dispensed with the many-one character of the pure form of cross-level realization. We saw that this move helped in the constructing of a version of Q-realization, but at the cost of collapsing the resulting account of realization into a form of same-level realization. In the next section, we will look at Melnyk’s version of cross-level realization. Melnyk does not insist on the many-one character of cross-level realization, and to that extent, like Shoemaker, is in danger of seeing his account collapse into a standard same-level account of realization. To put this in terms he does not use, but that I think gets at the crux of the matter, he prevents Shoemaker’s collapse by insisting that his version of realization *just is* cross-level. This has the effect of making his version of cross-level realization, *brute necessitation*. But that results in a relation that is neither X-realization nor Q-realization, neither exoneration nor robustly explanatory. Invocations of Melnyk’s realization are merely insistences that the world is *like this*. 
§5

§5.1 We saw in §4.4 that Andrew Melnyk criticized Sydney Shoemaker’s version of cross-level realization -- microrealization -- on the grounds that it did not posit a “substantial metaphysical relationship” between realizer and realized property instances. According to Melnyk, although it is true that microrealized instances occur just in case microrealizer instances of an appropriate type occur, Shoemaker’s microrealization does not explain this co-occurrence physicalistically, or indeed at all. To put this in the terms I am using in this chapter, Melnyk’s charge is that instances of Shoemaker’s microrealization are not even instances of X-realization. Realized property instances might, for all Shoemaker’s microrealization demands, be instances of emergent properties that just so happen to co-occur with instances of their physical realizers, and so there is no guarantee that microrealized property instances are physicalistically exonerated by their realizers. (Recall that for x to “exonerate” y is for there to be some other relation between x and y that explains y’s existence in a legitimate way). Here is a third, more traditional way to put Melnyk’s charge: Melnyk claims that microrealization is not a form of superdupervenience.\footnote{See Horgan (1993). Any relation between entities x and y that both entails the strong supervenience of y on x and is such that its existence (i.e. the relation’s existence) may be explained by appeal to explanatory resources found at x’s level of nature is a form of superdupervenience. Realization is supposed to be a form of superdupervenience.}

I belabor the point about Melnyk’s charge against Shoemaker’s microrealization because I want now to turn to Melnyk’s own account of realization, and it is not obvious that, or how, it escapes a charge that is quite similar to the charge that Melnyk levels against microrealization. As we will see, Melnyk’s account of realization does posit a “substantial metaphysical relationship” between realizer and realized properties, but it is not clear that this relationship is satisfactorily explanatory. The relationship looks, in fact, like brute necessitation.
Part of the larger to point to be drawn from an analysis of Melnyk’s account of realization is that if an account of realization that can be shown to identify realization with brute necessitation -- whether the identification is explicit or tacit, intended or not -- then the relation described by that account fails to qualify as superdupervenience. The invocation of a relation that fails to count as a species of superdupervenience will, in turn, fail to explain the occurrence of the realized property in terms provided by the explanatory resources to be found at the level of nature at which the realizer occurs.

Here is another, perhaps clearer way to express the worry. We saw previously that accounts of realization owe us accounts of a relation that, when invoked, explains why it is that the object that has the realized property has this property rather than that property. If properties are distinguished by their causal powers, then what the account of realization owes us is an account of why it is that the realizer property or properties bestow the object with the realized property these rather than those causal powers. In cases of same-level realization this is relatively easy to achieve because one and the same object has both the realizer and realized properties (so there is no mystery about how the bestowing works) and the causal powers are the same for both properties. In cases of cross-level realization that adhere to the characteristics laid out in Table 1, however, it is more difficult.

In the case of Gillett’s dimensioned realization, we saw that the relation fails to explain why the realizer properties and relations bestow these rather than those causal powers upon the object with the realized property. Each case of dimensioned realization is in need of filling in with explanatory details from the sciences. The explanation for why these causal powers rather than those are realized can only be given on a case-by-case basis.
In the case of Shoemaker’s microrealization, we had an example of a philosopher who was willing to sacrifice one of the three standard characteristics of cross-level realization in order to get a relation that was robustly explanatory. Shoemaker sacrificed the many-one character of cross-level realization; we saw that when that goes, the requirement of qualitative difference between the realized and realizer properties goes with it. With those characteristics of cross-level realization eliminated, it was possible for Shoemaker to build something very much like Jaegwon Kim’s Causal Inheritance Principle into his account of cross-level realization. The Principle is implicitly doing the heavy lifting of explaining why the object with the microrealized property has these rather than those causal powers.

But there is another way to create a version of cross-level realization that ensures, when invoked, that these realizer properties will result in specific causal powers being bestowed upon the object with the realizer property -- a way that does not require any case-by-case filling in by the sciences and that will work even if the the causal powers of the realizer properties are qualitatively different than the causal powers of the realized property. And that is simply to explicate cross-level realization as brute necessitation. We can try saying that it is just a brute fact that when these realizer properties are instantiated, those other realized properties (however different from the realizer they may be) will be instantiated.

My criticism of Melnyk’s account of realization is that it implicitly explicates realization as brute necessitation.

§5.2 Melnyk provides his account of realization, and an account of the physicalistic picture of the world that it underwrites, in his 2003 book A Physicalist Manifesto. In §5.3 I will discuss Melnyk’s formal definition of realization as provided in that book, showing that it ultimately
explicates realization as brute necessitation. First though, I would like to prepare the way by
discussing an informal but concise description of Melnyk’s account of realization that he
provides in a later, 2006 paper. I will do that in this section; then, in the next section, §5.3, I will
discuss the version of physicalism that Melnyk’s account of realization is intended to underwrite.
Once these tasks are done, it will be easier to see the importance of the details of Melnyk’s
formal definition that I wish to highlight in §5.4.

Here is the informal characterization of realization from Melnyk’s 2006 paper.

[R]ealization is a relation that holds not between types (e.g., between properties or
event-types) but between tokens of types (e.g., property-instances or event-
tokens). Moreover, a realized token can be realized only if it’s a token of a
functional type. Note, however, that the account uses ‘functional type’ very
liberally indeed, to refer to any type whose tokening just is the tokening of some
or other type that meets a specific associated condition, where this condition
could be of any kind, and needn’t be the playing of a causal role. Finally, a
realizing token realizes a token of a given functional type by being a token of
some or other type that meets the specific associated condition for that functional
type. (Melnyk, 2006, p. 129; italics original)

The primary point I will press about Melnyk’s account of realization is this: it seems to
me that the whole question of the viability of Melnyk’s account comes down to just what he
means when he says that a functional type is “any type whose tokening just is the tokening of
some or other type that meets a specific associated condition.” If this could be read as meaning
that realized tokens of functional types are identical to the tokens that are their realizers then
Melnyk’s account of realization would be, at least to this extent, in excellent shape. If we could
read the account this way, then realizer tokens would “make real” realized tokens by being the
realized tokens. We would have, in Melnyk’s account, a version of token physicalism
underwritten by a version of the realization relation, and all of this would seem to be
explanatorily unexceptionable. There would no mystery about how realized tokens of functional properties come into existence if they just are tokens of physical properties.\textsuperscript{17}

But that seems not to be what Melnyk means. When he says that “the tokening” of a functional type just is “the tokening” of some physical type, this seems not to be quite the same thing as saying that the token of one just is (= is identical to) the token of the other. Melnyk claims that for there to be a token of the realized type just is for there to be a token of the realizer type; but this seems not to be precisely the same claim as the claim as the claim that the realizer and realized tokens are identical (Melnyk, 2003, pp. 31, 51).

We know that Melnyk does not intend “just is” claims of these sorts to indicate token identity claims because, first of all, he explicitly denies that in cases of realization the realized token must be identical to the realizer token.\textsuperscript{18} But if the explicit denial is not sufficient, he further does not require that realized properties be instantiated in the same objects as their realizer properties (Melnyk, 2003, p. 21), so realizer and realized property instances could not be identical. Further still, Melnyk allows that not just tokens of properties but events and objects

\textsuperscript{17} This, for example, seems to be how Joseph Levine understands Melnyk’s book.

The basic idea of realization physicalism, then, is this. There is a fundamental level of reality, the physical. Above that are phenomena that are essentially characterizable in terms of roles of various sorts. There is in fact a hierarchy of levels of roles, some realizing, or implementing, others. But in the end, the ultimate role fillers are the fundamental objects, properties, and events of physics. (Levine, unpublished)

But this is a mistaken reading if Levine means that in Melnyk’s system there are only physical role-players and then a lot of upwardly-ascending roles. In Melnyk’s system it is not that straightforward. The conditions under which a functional type is tokened need not be the the playing of a causal role, and the functional token need not be identical, or even co-located with, the token of its role-filler (or, more broadly, its condition-meeter). So while it is true in a sense that the ultimate role-fillers are physical, that is true only in the sense that it is only physical objects, properties, and events, that do not need to be necessitated into existence by other objects, properties, and events. The character of the necessitation is what is at issue here. Levine seems to think it is identity, but Melnyk does not require that -- indeed, he denies that it is identity without saying much else about it. That is why I argue that in Melnyk’s system the non-physical tokens are brutally necessitated into existence by, ultimately, the physical tokens.

\textsuperscript{18} Of his “canonical formulation” of realization, Melnyk writes, “it does not entail that if token x realizes token y, then token x = token y. As far as this definition goes, it is an open question whether a realized token is identical with its realizer” (Melnyk, 2003, p. 21). We will get to the canonical definition itself shortly.
may be realized (provided that they are tokens of “functional” event and object types) . . . but he does not require that realized objects have objects as realizers, or that realized events have events as realizers, or that realized properties have properties have realizers. Rather, he allows that realized entities of any one of these three ontological kinds may have as realizers entities of any one of these three ontological kinds. Events can realize objects, property instances can realize events, and so on (Melnyk, 2003, p. 31). So we have many reasons to read Melnyk’s frequent claims that the tokening of a functional type “just is” the tokening of a physical type to be something other than claims of token identity (the most compelling reason being the explicit denial, of course).

But then, the explanatory adequacy of Melynq’s realization relation rests on just what he means by “just is,” because the “just is” is doing all of the explanatory work in his account of physicalism. I will develop this point in the next subsection.

§5.3 Melnyk describes his version of physicalism -- “realization physicalism” -- as a “retentive physicalism” because it allows us to retain all of the special- and honorary-scientific kinds that we use in everyday discourse (Melnyk, 2003, p. 32). Although on his account everything is ultimately explained in terms of the physical, we are allowed to continue to believe in the existence of such non-narrowly-physical types as tectonic plates and currency, football games and finger snaps, because such types of things may be identified with functional types, and the

---

19 Melnyk writes:

[L]et us call a version of physicalism retentive if, and to the extent that, it does not require denying the existence of tokens of special-scientific or honorary-scientific types -- those types which are spoken of as such in the special sciences or in the honorary sciences such as folk psychology, folk physics, and other bodies of folk knowledge developed historically as various parts of arts and crafts. (Melnyk, 2003, p. 32)

In a footnote Melnyk adds “If commonsense moral claims should turn out to be knowledge, then moral discourse would also qualify as an honorary science” (p. 32, f. 29).
existence of tokens of functional types is guaranteed by the existence of tokens of physical types. For each functional type there is an associated condition, and for there to be a token of a functional type just is for there to be a token of some physical type that meets that functional type’s associated condition. So we have in Melnyk’s realization physicalism three claims about the relationships between various kinds of things.

(a) Special or honorary-scientific types are identified with functional types.

(b) Tokens of physical types meet functional types’ associated conditions.

(c) For there to be tokens of a functional types “just is” for there to be tokens of physical types that meet the functional types’ associated conditions.

It is crucial to see that though (a) is an identity claim, (c) is not, as discussed in §5.1. It is (c) that does most of the work in characterizing the relation between the token of the physical type and the token of the functional type -- which is to say that it is (c) that does the most work in characterizing the realization relation. Despite this, Melnyk spends most of his time explaining (a) and (b). In the remainder of this section I will unpack (a) and (b). The unpacking of (c) will constitute an unpacking and critique of Melnyk’s account of realization; this I do in §5.4.

(a) Under Melnyk’s realization physicalism it is a necessary condition for the license to retain in our ontology tokens of special and honorary scientific types that such types be identified with functional types. A functional type is a type that is tokened just in case some token exists that meets that type’s associated condition -- so, at minimum, a functional type is one that is definable via an associated condition. Melnyk writes that it is an a posteriori matter whether or

\[20\] That is the aspect of Melnyk’s work that makes it difficult to understand at times, it seems to me. It is very easy to read that special and honorary scientific types are identical to functional types and then forget that tokens of physical types are not necessarily identical to tokens of functional types -- despite the frequent claim that the tokening of one just is the tokening of the other.
not a given special or honorary scientific type is identical to a functional type. In order to
determine whether or not a special or honorary scientific type that we wish to retain may be
retained, one of the things we have to do is go out into the world, do empirical science, and
compare the type we hope to retain to the extant functional types, to see if we can get a match.\footnote{By “extant functional type” I mean a functional type that both exists and is tokened in the actual world.}

This is a very unusual way of talking about functional types and their relation to special
and honorary scientific types. It is much more common to say that \textit{functional reductions} of
special and honorary science types proceed by first “functionalizing” the special or honorary
scientific type in question and then looking for the physical property instances (or “tokens”) that
play the roles thus determined to be definitive of the functionalized honorary or special scientific
type.

The method by which a special or honorary scientific type is “functionalized” is usually
construed as \textit{a priori}, a matter of conceptual analysis whereby we discover to what functional
type an ordinary special science term refers. Levine calls this process “a (relatively? quasi?) \textit{a}
priori process of working the concept of the property to be reduced ‘into shape’ for
reduction” (Levine, 1993, p. 132). Lewis, Kim, and others would take it to be a matter of
definition, or nearly so, just which causal role, and therefore just which functional kind, each
special science kind \textit{is} (Kim, 1998, p. 24; Lewis, 1972).

By contrast, Melnyk writes as if the basic sciences reveal to us a world of functional
types, and what we have to do if we want to preserve (or “retain”) tokens of special or honorary
scientific types is determine though empirical investigation whether or not our favored special-
or honorary-scientific types are identical to one of the extant functional types. Note, this is \textit{not} a
reductionist picture. The idea is not to find in the world through empirical investigation the physical tokens that meet the functional conditions associated with a given functional type and then to identify the tokens of special or honorary-scientific types with the condition-meeting tokens of physical types. Rather, the idea is that we have already found (or anyway are capable of finding, independently of any desire to retain tokens of special or honorary scientific types), through empirical investigation, tokens of functional types that may or may not be identical to the tokens of physical types. Given that such functional tokens exist, the functional types exist and are tokened in the actual world. Functional types that meet these conditions are available for identification with honorary or special scientific types. It is then up to further empirical investigation to determine whether such identification is warranted.

Here’s Melnyk describing the process.

Statements reporting the requisite identities, if discoverable at all, will only be discoverable a posteriori, nondemonstratively inferred from empirical premises, and even then perhaps only with great difficulty. Here is one way . . . First, by observation, we might discover that special-science type, P, is tokened in a given location when and only when functional type, F, is tokened in that same location; then, since the hypothesis that P=F would, if true, arguably provide the best explanation for the perfect coincidence (within our experience) of P-tokens and F-tokens, we could argue by inference to the best explanation from our observational finding to the conclusion that in fact P=F . . . (Melnyk, 2003, pp. 34-35)

Nowhere does Melnyk explain what it would be to observe a token of a functional type as opposed to or in the absence of a token of a special- or honorary-scientific type. He does not explain what it would be to observe, as it were, a token of a “bare functional type.” There are two points to make about this. First, he owes us such an account, since his version of physicalism hangs from its viability. Second, to repeat, this identification is not the identification of tokens of functional types with tokens of physical types. That is a completely separate matter.
and at any rate, according to Melynk, it is not necessary that tokens of functional types be identical to tokens of physical types in order for (tokens of) physical types to realize (tokens of) functional types.

(b) Now we turn to (b) and the question of what it is for a token of a physical type to meet a functional type’s associated condition. We are still not, in this subsection, asking after the relationship between tokens of physical types and the tokens that they realize, of functional types -- which is to say that we are not, in this subsection, asking after the realization relation. We are simply asking after what it is for a physical token to meet a condition.

Melynk is clear about the following: that a physical token meets a given condition must follow logically from the physical facts alone. That is, given a description of the physical token and the “very broad realizer” (2003, p. 29) that consists in the physical token together with whatever other physical facts are relevant to the meeting of the condition, it must be possible to deduce that the condition is met. Melynk writes,

But if a physical type’s meeting condition C must follow logically from the physical facts alone, then condition C cannot just be characterized in any old way, else there will be no valid derivation, even in principle, of the conclusion that C has been met from premises that specify purely physical facts; indeed, for the derivation to be valid, condition C must be characterized in terms that are either physical or quasi-logical or a mixture of both. (Melynk, 2003, p. 25; italics original)

On standard causal role functionalist views, the “condition” is simply a causal role; to “meet the condition” is to “play the role.” A causal role is a set of causal relations that a realizer token must be able to enter into. To put it somewhat awkwardly, causal roles are neither temporally nor spatially distant from the tokens that play them; to put it less awkwardly, on
standard accounts, whether a token plays a role may be determined by checking facts that are both strictly causal and purely local to the token itself in both time and space.

But on Melnyk’s system a much wider variety of types of condition may characterize a functional type; a much wider variety of conditions may be the sorts of conditions that a physical token must meet in order to realize a functional token. I quote at some length:

One familiar kind of associated condition, of course, is that of playing a certain causal role: to meet the condition, a type must be such that its tokens cause and are caused in certain, specified ways. But an obvious variation would be the playing of a noncausal nomic role: to meet a condition of this sort, a type must be such that, as a matter of law, it is tokened iff (but not because) some other, specified type is tokened.

The laws invoked in specifying an associated condition, then, might be causal or noncausal; and we can add that laws of either kind need not be strict and deterministic, but might also be probabilistic or hold only ceteris paribus (however that should be understood best). Similarly, a lower-order type’s meeting of an associated condition might be a matter of its tokens’ standing in other relations that nomic ones (e.g. spatiotemporal relations). Again, a type’s meeting an associated condition might require the holding of certain, specified circumstances that are external to, and/or earlier (or indeed later) than, tokens of the that type . . . so special- or honorary-scientific types whose tokens do not supervene upon simultaneous and local physical conditions might still be functional types in my sense. Biological types, for instance, might be functional types, even though, because they are sometimes defined in terms of a Darwinian selectional history of a certain sort, an atom-for-atom duplicate of (say) my heart that formed by accident in interstellar space would not be a heart. (Melnyk, 2003, pp. 38-39)

I provide a long quote in order to give the scope of the idea. Conditions of those sorts are the conditions that must be specified in strictly physical terms, and that a physical token meets such conditions must be derivable from a physical description of a physical token and its very broad realizer.

I want to pause here to point out another oddity of Melnyk’s account of realization physicalism. On most accounts of physicalism, or of functional reduction generally, it is a
relatively a priori matter whether a special science type may be characterized by a causal role or other associated condition, and it is a strictly empirical matter, discovering which if any physical or lower-level tokens in the world play the causal roles or meet the associated conditions of the functional types. Melnyk has this reversed. On his account it is an a posteriori matter of empirical investigation whether a special science type is (identical to a) functional type, and a purely logical matter whether a physical token in fact meets the functional type’s associated condition. On most accounts it would be enough to 
*look and see* whether a physical token is playing the requisite causal role (or whatever), but Melnyk demands a *deductive proof* that it is, a proof written in the language of physics.

Thus if we consider Gillett’s account of cross-level realization, it is experiment and not deduction that determines whether protein subunits behave in ways appropriate to the realizing of the property of being a voltage sensitive gate -- a property that is itself determined to be functional and to have such-and-so an associated condition through relatively a priori (not to say merely *obvious*) means. Melnyk would say, contrariwise, that it is a matter of investigation to determine just which functional type it is that the property of being a voltage sensitive gate is identical to, and a matter of deduction to establish whether the protein subunits meet that functional type’s associated condition. True, empirical research will be needed to provide the complete physical description of the neural subunits and the very broad realizer that facilitates their activity, so that we may then perform a deduction that uses that description as a premise, but it is still the deduction, not the research, that is bringing in the wash. Thus Gillett’s and Melnyk’s takes on these matters are opposite to each other, and I take it that, to this extent at least, it is Gillett that is making use of the standard view.
§5.4 It is time now to address (c) directly, the claim about the relation that obtains between the tokens of physical types that meet C and the token of functional types that are “realized” by those physical tokens. Since this relation, whatever it turns out to be, is what Melnyk is telling us is the realization relation, this is a good place to relate his formal definition of the realization relation.

Token $x$ realizes token $y$ iff (i) $y$ is a token of some functional type, $F$, such that, necessarily, $F$ is tokened iff there is a token of some or other type that meets condition, $C$; (ii) $x$ is a token of some type that in fact meets $C$; and (iii) the token of $F$ whose existence is logically guaranteed by the holding of condition (ii) is numerically identical with $y$. (Melnyk, 2003, p. 21)

My claim is that Melnyk leaves realization unexplained. He explains everything else -- what it is for a realizer to meet a condition and what it is for a special or honorary scientific type to be identical to a functional type. But he does not explain how it is that the token of the physical type necessitates the existence of the token of the functional type. In the formal definition the necessity is invoked in clause (i). $y$, a token of a functional type $F$, “necessarily” exists if and only if (among other things) “there is a token of some type or other that meets $C$.” But since $y$ is not identical to the token, $x$, of the type that meets $C$, some further explanation of what the relation between $x$ and $y$ is, is needed. In clause (iii) it looks as though a further explanation is on offer: the existence of $y$ is “logically guaranteed” by the fact that $C$ is met by $x$. But this appearance is a bit deceiving. What is “logically guaranteed” is that $x$ meets $C$, given a description of $x$ and its broad realizer. (As discussed in §5.3, (b)) What is being said in clause (iii) of the formal definition is that given that a token of $F$ necessarily exists when $C$ is met, and given that the fact that $C$ is met is logically guaranteed by the description of $x$ and its very broad realizer, it follows in a harmlessly extended sense of “logical guarantee” that it is logically
guaranteed that a token of F exists. But obviously this does nothing to explain the necessity, *itself*, of x’s existence, given that C is met.

I spent the past fews pages discussing the a posteriori identity of special- and honorary-scientific types with functional types, and the relation of “meeting” that holds between a token of physical type and a condition, so that I could keep them clearly separate from the relation that really matters, here, and that is the relation between the physical token and the functional token that it “realizes.” The primary point I wish to make is that despite all of the work that Melnyk does to explain what goes into identifying a special- or honorary-scientific type with a functional type, and into specifying functional types’ associated conditions, and how to meet those conditions, he does nothing to tell us what the necessitation relation is between x and y. The closest that he comes is in chapter two -- the chapter to which he directs us for an account of the “musts” in this passage, from chapter one:

Given the existence of a token of a type that meets C, there **must** exist a token of F, since for there to be a token of F just *is* for there to be a token of some type that meets C. Moreover, by the definition of “realized,” *this* token of F is one and the same as the F-token we were talking about in the first place. So, given the physical facts mentioned, the F-token we were talking about in the first place **must** exist. (Melnyk, 2003, p. 31; italics original; boldface added)

This passage ends with a footnote that reads, “In the next chapter [i.e. chapter two], we look more closely at the nature of these very strong **musts**” (pp. 31, f. 28; italics original). So it is fair to say that if no explanation of the “necessarily” in clause (i) of the formal definition is to be found in chapter two then it will not be found in the book. And no explanation is to be found in chapter two.
Although Melnyk does not provide a page number for where in chapter two we should look for the discussion of the “musts,” I think there are only two passages that he could be talking about. Here is the first:

As we saw in chapter 1, realization physicalism entails that, for every nonphysical token that falls with the scope of (R)\(^{22}\), there are physical conditions -- the token’s very broad realizer -- that necessitate the existence of that token. For consider some particular nonphysical but physically realized token that falls within the scope of (R), and suppose that it is a token of a nonphysical type F, with associated condition C; then, given the existence of a physical narrow realizer, the holding of certain physical laws, and the holding of any external or historical physical conditions that are relevant, there must exist a token of a type that meets C; but given the existence of a token of type that meets C, there must exist a token of F, since for there to be a token of F just is for there to be a token of some type that meets C. (Melnyk, 2003, p. 51; italics original)

This is just a restatement of the *must*, not an explanation of it. We are no closer to understanding why there must be a token of F given that there is a token of a physical type that meets C. The second passage in chapter 2 to which Melnyk could be referring when he says to look to chapter 2 for an accounting of the “musts” in chapter 1 is this:

A realizationist can also provide an attractive account of the *necessity* recorded in (M)\(^{23}\). To see how, let “P” be a complete physical description of the actual world (not including the physical laws), let “L” express all the laws of physics that hold in the actual world, and let “N” be a statement, framed in the proprietary

\(^{22}\)“(R)” is not the formal definition of realization but a statement about the ontology of the world, given the truth of realization physicalism. Here it is:

(R) Every property instance is *either* an instance of a physical property *or* a physically realized instance of some functional property; every object is *either* an object of some physical kind *or* a physically realized object of some functional object kind; every event is *either* an event of some physical event kind *or* a physically realized event of some functional event kind. (Melnyk, 2003, p. 26; italics original)

\(^{23}\)Here is “(M)”:  

(M) Any possible world that (a) has exactly the same distribution of physical tokens as does the actual world and in which (b) exactly the same laws of physics hold as hold in the actual world contains all the tokens of nonphysical types, falling within the scope of (R), that the actual world contains. (Melnyk, 2003, p. 51)
vocabularies of the special and honorary sciences, that asserts the existence of all those tokens of nonphysical types, falling within the scope of (R), that actually exist. Then (M) entails that:

\[(M^*) \text{ Necessarily, if } P \text{ and } L, \text{ then } N.\]

The necessity operator here expresses the strongest sort of necessity, since it has to express the idea, made explicit in (M), that \emph{any possible world whatever} in which it is true that \(P\) and \(L\) is one in which it is also true that \(N\). But what, if anything, \emph{explains} the strong necessity of “If \(P\) and \(L\), then \(N\)”?

This is the key moment: what \emph{explains} the necessity of the existence of the non-physical tokens? The answer to this question will also answer the question: what is realization? Melnyk continues:

Realizationism answers that the conditional’s strong necessity need not not be treated as a brute modal fact, for it has an explanation of the following sort.

(So, just there, Melnyk denies that realization is brute necessitation; that is, he denies exactly what I am accusing him of. He goes on to explain why realization is not brute necessitation:)

From the claim that \(P\) and \(L\), it follows, via rules of natural deduction, together with analytic trivialities of the sort mentioned in the previous chapter (e.g. “If something has charge, then it has at least one property”), that there exist many tokens of certain functional types whose associated conditions are specifiable in physical or quasi-logical terms . . . (Melnyk, 2003, p. 53; italics original)

But this is an error. What “follows via the rules of natural deduction and analytic trivialities” is that the conditions are met. What does not follow is the fact that the conditions are associated with functional types, or that meeting the conditions would necessitate the existence of tokens of functional types. Once again, \emph{given the associations}, that is, given that the existence of tokens of certain functional types is necessitated by the meeting of specified conditions, it follows from \(P\) and \(L\) that such functional tokens exist -- but \emph{that} is the necessitation that \emph{is} realization and \emph{that} is the necessitation that needs explaining. This is the necessitation for which
Melnyk promised an explanation in footnote 29 of chapter 1, but it is not quite the explanation that he actually provides. Melnyk explains (b), not (c).24

§5.5 In this section, §5, we have looked at an attempt to provide a notion of realization that is very multi-purpose. It is cross-level to the extent that it is supposed to work for a great many kinds of explanation, many of which may be construed as cross-level. But what we are left with is a relation that is not clearly Q-realization and not clearly X-realization. It is not clearly Q-realization (indeed, it is clearly not Q-realization) because it is not robustly explanatory of the presence of tokens of the realized functional type. To be sure, Melnyk provides us with an account of what it is to meet a certain kind of condition -- and he tells us that meeting various conditions of this kind just is to bring into the world tokens of functional types. But that just is what is doing the work of realization. Meeting the condition just does make real a token of the functional type. But we are not told how. That is why I claim the necessitation is merely brute.

And calling brute necessitation “realization” does not make it less brute. Indeed, it goes the

24 For the sake of completeness I will continue directly with Melnyk’s text, because from where I left off it continues to look as if he is about to explain the key relation, (c), without ever quite doing it. Thus, continuing directly, we get an explanation of (a), not (c):

... let “F” assert their existence [i.e. the existence of the tokens of functional types], specifying exactly which tokens they are and of what types. Now “F” does not automatically entail that N, either analytically or via rules of natural deduction, because while “F” is expressed in physical or quasi-logical terms, “N” is by contrast expressed in the proprietary terms of the special and honorary sciences, and there are, I take it, no analytic connections between physical terms and those of the special and honorary sciences. However, retentive realizationism claims that every (tokened) special- and honorary-scientific type is identical with some or other functional type whose associated condition is specifiable in physical or quasi-logical terms; moreover, given realizationism, it is precisely these functional types whose tokening follows from the claim that P and L. But if, given the claim that P and L, if follows that F (i.e., that tokens of these functional types exist), and if these functional types just are those special- and honorary-scientific types asserted to be tokened by the claim that N, it follows intuitively that, given the claim that P and L, it must be true also that N. (Melnyk, 2003, p. 53; italics original)

Although this part of the passage is primarily an unpacking of (a), it alludes to the mistaken claim that F -- (the claim “that tokens of these functional types exist”) -- follows from P and L. Again, what follows from P and L (sentences that describe the distribution of tokens of physical types and the laws of physics) is just that certain conditions are met. That tokens of functional types necessarily exist given that the conditions are met does not follow from P and L or from anything else that Melnyk has discussed. To think so is to conflate (b) and (c).
other way around: explicating realization as brute necessitation takes away the explanatory point of positing a realization relation.

Melnyk’s realization is not X-realization, either, because to say that a token of a physical type meets an associated condition and thereby necessitates the existence of a token of a functional type is not to invite the sciences to “fill in the details” of the necessitation. There is no further relation that does the work of exoneration, an exoneration that is merely recorded by X-realization. Melnyk’s realization, unlike, say, Gillett’s, is not supposed to be like the bare exoneration relation. And indeed it is not. There is nothing to appeal to, no science, underneath brute necessitation. That is what makes it brute.
Chapter 6

A Brief Conclusion

§1

In this section I recap some of the work in the preceding dissertation, then in §2 I suggest lines of future research.

In this dissertation I have taken a look under the hood of the realization relation to see what makes it tick. My primary goal has been to critique the most common way of explicating the notion of realization: via the causal role-playing relation. My guiding suspicion has been that the causal role-playing relation cannot do the work that philosophers want to have done when they call upon realization. I believe that I have established that my suspicion is well-founded. The Master Argument I laid out in §8 of Part I of this dissertation (and defended in Part II) shows that none of the relata of the causal role-playing relation are suitable candidates for being the kind of property that philosophers are talking about, when they talk about realized properties. The notion of realization is supposed to solve a philosophical problem; the notion causal role-playing isn’t up to the task.

But let me state in more direct terms what I take myself to have shown. I have shown that two of the central tenets of the idea the realization is causal role-playing are unworkable. The first of these tenets is so central to the philosophical view underwritten by causal role realization that it is not typically noticed or given a name. I call it “the Principle of Causal Subveners.”
The Principle of Causal Subveners. For any instance of a realized property \( R \), the realizer property \( P \) upon which the instance of \( R \) strongly supervenes figures in the causal explanation of the effects of that instance of \( R \).

The Principle of Causal Subveners is held by reductive and and anti-reductive physicalists alike who appeal to causal role realization. In the non-reductive case, realized properties are taken to be identical to role properties, defined by GRP. In the reductive case, realized properties are taken to be identical to the properties upon which they strongly supervene -- that is, instances of realized properties are taken to be identical to the instances of their total realizers.¹

The second central tenant of causal role realization that I take myself to have refuted is the Causal Inheritance Principle. Unlike the Principle of Causal Subveners, the Causal Inheritance Principle has been explicitly noted and adopted by advocates of causal-role realization.

[The Causal Inheritance Principle] If mental property \( M \) is realized in a system at \( t \) in virtue of physical realization base \( P \), the causal powers of this instance of \( M \) are identical with [or are a subset of] the causal powers of \( P \). (Kim, 1993: 326; italics original)

Others have established that the \( P \) referred to by Kim when he states the Causal Inheritance Principle is the total, not the core, realizer of \( M \) (Antony and Levine, 1997, Witmer, 2003). Thus, the \( P \) mentioned in the Causal Inheritance Principle is the same sort of property as the \( P \) mentioned in the Principle of Causal Subveners. Given this, its easy to show that the correctness of the Principle of Causal Subveners is a necessary condition for the correctness of

¹ Arguably, the only philosopher who uses causal role realization to formulate a version of physicalism but who doesn’t subscribe to the Principle of Causal Subveners is David Lewis, since he identifies realized properties with their core realizers. I argue that this cannot be right because realized properties don’t even strongly supervene on core realizers, a necessary condition for identity. Lewis tries to deal with this problem in “Mad Pain and Martian Pain” (Lewis, 1983); unconvincingly, in my view -- and apparently in most philosophers’ views, since to my knowledge no one currently adopts Lewis’s solution.
the Causal Inheritance Principle. (To put that another way, it is easy to show that the Causal
Inheritance Principle implies the Principle of Causal Subveners.)

It should not be surprising that Causal Inheritance Principle implies the Principle of
Causal Subveners. The Causal Inheritance Principle says that a realized property has the same
causal powers as its total realizer. The Principle of Causal Subveners says that a realized
property’s total realizer has the same effects as the realized property itself. Modulo some details
(details which keep the two Principles from being strictly equivalent), the two principles say the
same thing but with different emphases.

The reason that both Principles fail, I have argued, is that total realizers are not causal
properties. I think that the main reason that philosophers have operated for so long under the
assumption that total realizers are just as much causal properties as are core realizers is that
philosophers have assumed that total realizers are simply core realizers plus causal background
conditions. But to think that is to confuse a property’s being implicated in a causing with a
property’s being implicated in a causal role-playing. Total realizers are not cores plus causal
background conditions; they are, as it were, cores plus causal-role background conditions. As
such, total realizers don’t do anything, causally. What they do is subvene role properties.

§2

It seems to me that my dissertation suggests two different avenues for future research.
Which line strikes the readers as more promising will depend, I suspect upon whether the reader
takes my work here to show that causal-role realization needs revision or abandonment. If one
wishes to preserve causal-role realization, one could try to develop a theory of mental states and
mental causation that is built on causal-role realization minus the Principle of Causal Subveners. Or, if one thinks that my arguments show that causal role realization ought to be junked, one could try to understand what is left of physicalism once causal role realization is left behind. Both projects strike me as worthwhile. Allow me to say something about each. I will say something about the first project in the remainder of this section, and the second project in the next section.

If we keep the doctrine that is definitive of non-reductive causal role functionalism, that mental properties are causal role properties of some sort, but abandon the Principle of Causal Subveners, we get a picture in which the kinds of mental states that are named in Lewisian psychological theories -- beliefs, desires, pains, perceptions, the definitive characteristic of each of which is its causal role -- exist but do not cause anything. We get, that is to say, a picture of epiphenomenalism about such mental states, but I believe it is possible that it is an epiphenomenalism that is worth pursuing -- an epiphenomenalism in name only.

What is at issue is a philosophical puzzle at the intersection of recent work in the metaphysics of mind and recent work in a newer philosophical field of inquiry that is sometimes called “the phenomenology of agency.” The puzzle at this intersection is: how do minds cause behavior? Philosophers working on realizationist metaphysics of mind over the past thirty or forty years -- the philosophers who have tinkered with functionalism and causal role realization, the philosophers whom I have been discussing in this dissertation -- have adopted a rather primitive view of the way in which minds cause behavior. They think of behavior as caused by functionally defined beliefs, desires, perceptions, and so on (we will see a classic statement of

---

2 For an overview see “The Phenomenology of Agency” by Tim Bayne (Bayne, 2008). I will be drawing from Bayne’s article in the following.
this simplistic view from Jerry Fodor, in a moment). Philosophers working in the more recent
field of phenomenology of agency have developed much more sophisticated and interesting work
on the relation between mind and behavior.

Causal role realizationists, working from the Lewisian model of functional systems, have
treated the causing of behavior by mental states as just like the causing of mental states by other
mental states, and just like the causing of mental states by sensory inputs. These states -- mental,
behavioral, sensory-input -- are all of them to be thought of as of more or less the same sort. The
primary difference between mental states and sensory or behavioral states is that sensory or
behavioral states are named by “O-terms” in the psychological Theory T while mental states are
named by “T-terms.” What is named, the mental states themselves, are not supposed to be
metaphysically or causally different than sensory or behavioral states. (That is, they are not
supposed to be metaphysically different except for the wrinkly that non-reductive realizationists
will hold that mental states are second-order states while sensory and behavioral states are first-
order states. But this is not supposed to be a causally important difference, for non-reductive
physicalists.)

When this metaphysical/causal homogeneity of mental and physical states is combined
with a rather primitive belief/desire model of psychology, according to which beliefs and desires
straightaway cause behaviors, we get a picture of the mind/behavior interface that looks like this,
from Jerry Fodor:

I’m not really convinced that it matters very much whether the
mental is physical; still less that it matters very much whether we
can prove that it is. Whereas, if it isn’t literally true that my
wanting is causally responsible for my reaching, and my itching is
causally responsible for my scratching, and my believing is
causally responsible for my saying . . . if none of that is literally
true, then practically everything I believe about anything is false and it's the end of the world. (Fodor, 1987: 77)

Here, Fodor says that it is more important to retain the causal status of mental states than it is to remain a non-reductivist. What Fodor seems to think must be saved in order to thwart epiphenomenalism is a very specific form of belief/desire psychology according to which mental states such as “wantings” “literally” cause reachings, and mental states such as “believings” “literally” cause mouths to speak, and so on. Fodor is not an outlier in this regard: my sense of the functionalist literature is that the metaphysicians of mind currently working on the problem of mental causation take this simplistic view of the mind/behavior interface generally.

Philosophers working in the phenomenology of agency -- the attempt to understand the structure and content of “agentic experiences” -- do not take the sort of view sketched in the above quote by Fodor seriously (Bayne, 2008). According to philosophers working on the phenomenology of agency what we experience as agents is that intentions to act cause behavior, not the beliefs and desires that may lead to intentions. Or, to be more precise, what is up for debate in the field of phenomenology of agency is whether agentic experience (the name for the occurrent experience of raising one’s arm now, for example) causes behavior, or whether agentic experience records the causing of behavior by intentions. That is: whether occurrent agentic experiences consist in intentions, or merely record intentions (Bayne, 2008: 191). For our purposes here, what matters is that none of the philosophers in this debate take seriously the idea that in non-pathological cases we experience mental states other than intentions (for example, Fodor’s “wantings” and “believings”) as directly causing behavior. I quote at length:

According to what we might call ‘the causal account’, the experience of purposive goal-directed behaviour involves (or perhaps just is) an experience of mental causation. In its simplest incarnation, the causal account says simply that to
experience one’s movements as one’s actions involves experiencing them as caused by mental states (that one also experiences as one’s own). This sketch can be embellished in various ways. For example, one might suggest that the experience of goal-directed agency involves the experience of causal relations between mental states. Consider again the vignette with which I opened this paper, in which you experience yourself as opening a door in order to leave your apartment and thus escape a fire. Arguably, this scenario includes an experience of causal relations as nested within each other: you experience your intention to flee the fire as causing your intention to leave the apartment, and in turn experience your intention to leave the apartment as causing your intention to open the door. If one accepts that the basic experience of agency includes an experience of one’s intentions as causing one’s movements, then there is at least a prima facie case for thinking that one can also experience some mental states as causing other mental states.

But what should we say about the basic causal proposal, namely, that the phenomenology of agency involves the experience of one’s movements as caused by one’s intentions? Some theorists reject the causal account on the grounds that causation cannot be — or at least *is not* — experienced. I don’t have much sympathy with this criticism, for it seems evident to me that causal relations can be experientially encoded. (One sees the bat as *causing* the ball to ricochet away.) A more plausible objection holds that although causal relations can be experientially encoded, they are not contained within the normal experience of autonomous agency. (Bayne, 2008: 191-92)

The alternative to the suggestion that intentions cause purposive behaviors is the suggestion that *selves* cause purposive behaviors (Bayne, 2008: 194). Fodor’s suggestion, that desires and beliefs cause behaviors, is not considered, except as perhaps correct of pathological cases.

Perhaps the difference between the normal case and the pathological case is that in the normal case one experiences one’s *intentions* as causally efficacious whereas in the pathological case one experiences one’s *desires* (or emotions or urges) as causally efficacious. (Bayne, 2008: 192)

The phenomenology of agency is the attempt to understand the structure and content of “agentive experiences,” the experience of causing behavior as an agent, that is, the experience of
engaging in action. It the rough veracity of this sort of experience that workers on the metaphysics of causation are attempting to save.

My contribution is to point out that the work done by realizationists in the metaphysics of causation, to the extent that it can be saved, shows that what cannot be right is that mental states with Shoemakerian total realizers cannot be what cause behaviors, and therefore that veridical agentive experience cannot be recording that they cause behaviors. There is much work here to be done at understanding what kind of picture of mind develops when we combine the work in the metaphysics of mind with the work on the phenomenology of agentic experience.

§3

I don’t really believe that the positive proposal mentioned in the previous section is very promising. I think that if we are going to make headway in the philosophy of mind, particularly in the metaphysics of mind, it is not going to be by salvaging what is left of causal role realization -- either to aide in debates in the phenomenology of agency or anywhere else. Rather, the necessary work is in the field of metaphysics of mind itself, where a lot of foundational work needs to be rethought. This work is both harder and more invigorating than the work suggested in §2; it involves abandoning ideas developed in the previous century to make room for ideas in the new century.

What we must do, I believe, is go back to the origins of the idea of realization to try to understand what kind of metaphysical lacuna realization was supposed to fill in. This involves going all the way back to Ernst Nagel’s 1954 Presidential Address to the APA, to understand better what kind positive metaphysical proposal was in the air that became so fatefully attached
to the names “realization” and “naturalism.” What we must do, then, is decouple naturalism from the commitments that make the positing of realization seem necessary. The project would be to save naturalism from realization.

It seems to me that “realization” is in the first place, before it is explicated, the name of the embarrassing remainder of certain forms of physicalism that became popular in the second half of the twentieth century. These forms of physicalism became popular because the logical space of options in the the twentieth century made these forms of physicalism seem the best choices. My own suspicion is that realization became popular because it allows one specifically to tie together physicalism and behaviorism -- two “isms” that were very popular in middle and late twentieth century anglo-american philosophy.

But in any case, if we are to have some idea of what the prospects for “non-reductive physicalism” are in the new century we have to get a grip on what it can be without causal role realization, which is to say without the pervasive quasi-behaviorism that unfortunately infected everything it touched in 1950s and after. My dissertation has above all been aimed at showing the kinds of incoherence that result when one tries to get behavioral properties to do things, when one tries to have one’s physicalist cake and non-reductively eat it, too, with the help of elements of a reified and lingering behaviorism that go by the name of “causal role properties.”

This suggests an historical project, a critical project, and a constructive project. The historical project is to get a better picture of what was going on in the middle of the twentieth century that left philosophers wanting what the realization relation was advertised to offer. The critical project is to compare the philosophical lay of the land at the middle of the twentieth century to that at the beginning of the twenty-first, so that we may engage in a comparison of the
two and thereby a reevaluation of the kind of mind-body relation that we think we are obligated
to be after. The question is: what happens to physicalism (or anyway naturalism) once the
hangover from logical behaviorism (which is what I contend the idea of a “causal role” amounts
to) is swept away? The constructive project is to provide an alternative to currently-entrenched ideas about realization and the picture of mind and world that realization underwrites. The first step in this constructive project is to engage with recent work on ontological dependence relations to see what the current options really are for the mind-body relation. The work on dependence relations is incomplete but exciting, in that it provides us with a way to get a clearer view of the ontological options that we have, away from the haze that surrounds realization-talk.
Bibliography


